Year : 2017

Characterisation of computed tomography devices and optimisation of clinical protocols based on mathematical observers

Racine Damien

UNIL | Université de Lausanne

Institut de Radiophysique

Characterisation of computed tomography devices and optimisation of clinical protocols based on mathematical observers

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

Par

Damien RACINE

M.Sc. en physique
Université Joseph Fourier, Grenoble I - FRANCE

Jury de thèse

Président du comité de thèse: Prof. Laurent DECOSTERD (CHUV-UNIL - SUISSE)
Directeur de thèse: Prof. Francis R. VERDUN (CHUV-UNIL - SUISSE)
Expert: Prof. Hilde BOSMANS (KU Leuven, Leuven - BELGIQUE)
Expert: Dr. Yannick ARNOUD (LPSC, Grenoble - FRANCE)
Expert: Dr. Fabio BECCE (CHUV, Lausanne - SUISSE)

Lausanne 2017

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président· e	Monsieur	Prof.	Laurent	**Decosterd**
Directeur· rice de thèse	Monsieur	Prof.	Francis	**Verdun**
Experts· es	Monsieur	Dr	Yannick	**Arnoud**
	Monsieur	Dr	Fabio	**Becce**
	Madame	Prof.	Hilde	**Bosmans**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Damien Racine

Master ès sciences physique de l' Université Joseph Fourier / Grenoble I

intitulée

Characterisation of computed tomography devices and optimisation of clinical protocols based on mathematical observers

Lausanne, le 8 décembre 2017

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Laurent Decosterd

« Un pessimiste voit la difficulté dans chaque opportunité, un optimiste voit l'opportunité dans chaque difficulté. »

Winston Churchill

Contents

Remerciements

Il est naturel de remercier à la fin d’un tel travail tous ceux qui, plus ou moins directement, ont participé à leur manière à la réalisation de cette thèse.

Je tiens tout d’abord à remercier mon directeur de thèse, le Professeur Francis R. Verdun de m’avoir accueilli au sein du groupe d’imagerie médicale de l’Institut de radiophysique et de m’avoir témoigné sa confiance pour la réalisation de ce projet. Ses remarques judicieuses tout en me laissant une grande liberté et sa vision plus globale m’ont été d’une grande aide, me permettant de m’épanouir tout au long de ce travail et ont permis son aboutissement dans les meilleures conditions.

Je tiens également à remercier les membres de mon jury de thèse, qui ont accepté de dédier une partie de leur temps et de s’impliquer dans les différentes évaluations.

Mes remerciements sont également destinés à toutes les personnes qui ont été co-auteur de mes publications. La rédaction d’un article scientifique est un exercice difficile, et l’apport de différents regards, parfois éloignés, est toujours une richesse pour le débat scientifique. Parmi eux, j’adresse à Pascal Monnin et au Professeur François Bochud un remerciement particulier. A Pascal pour ses conseils et nos discussions concernant l’évaluation de la qualité d’images via les méthodes usuelles en traitement du signal et à François pour m’avoir accueilli au sein de l’Institut et pour ses explications et son expertise sur tous les sujets relatifs aux modèles d’observateurs. Sans oublier Anais, Alexandre, Georg, Julien, Nick pour leurs conseils concernant mes travaux, mais également pour tous les autres moments.

Merci également à Julie Bize, Christel Elandoy, Silvano Gnesin mes collègues du GIM pour leur bonne humeur au quotidien et l’ambiance de travail agréable. Un remerciement particulier à Christel et Julie, leurs regards de techniciennes du terrain m’a permis de garder les pieds dans l’application clinique, permettant à mon message de rester pertinent.

Je voudrais également remercier mes collègues à l’Institut qui m’ont permis de profiter d’une atmosphère plus qu’agréable durant les heures de travail (un remerciement particulier à Marie pour son assiduité lors de nos séances de course à pied à la mi-journée).

Je tiens aussi à remercier profondément les docteurs Hugues Brat, Fabio Becce, David Rotzinger et Sabine Schmidt pour nos diverses collaborations et pour leurs explications sur les aspects médicaux en rapport avec ce travail. Tous nos projets communs ont apporté une grande plus-value au travail accompli.

J’aimerais remercier toute ma famille et mes amis pour leur soutien durant ces trois années de thèse et plus généralement durant toute ma scolarité. Je pense tout d’abord à mes parents sans qui l’enfant que j’étais ne serait pas devenu l’homme que je suis. C’est avec émotion que je leur dévoile le fruit de mes efforts.

Enfin mes derniers remerciements et non les moindres, s’adressent à ma femme Anne-Lise, qui, pour mon plus grand bonheur partage ma vie et mes expériences professionnelles depuis leurs origines. Elle est le pilier de toutes mes constructions et la base de tous mes projets. Elle a su, tout au long de cette thèse m’encourager dans ma voie, son soutien a été sans faille. Elle est la clef de ma réussite, sans elle à mes côtés, cette réalisation n’aurait pas la même saveur.

Ce travail a été partiellement financé par le Bundesamt für Strahlenschutz (BfS), No. 320030-140995.

Résumé

Les évolutions technologiques des modalités diagnostiques d'imagerie par rayons X permettent aux radiologues d'améliorer la qualité du diagnostic et les soins aux patients. Dans ce contexte, le nombre d'examens radiologiques effectué en radiographie conventionnelle, fluoroscopie ou tomodensitométrie (TDM) est en constante augmentation. L'imagerie TDM contribue à environ 70% de la dose efficace annuelle totale délivrée à la population par l'imagerie par rayons X. Comme l'utilisation des rayons X en imagerie médicale est liée à un risque d'induction de cancer, risque décrit par le modèle linéaire sans seuil, développé traditionnellement pour la radioprotection des patients ; de nombreux efforts ont été mis en œuvre pour réduire l'exposition du patient afin de s'assurer que le bénéfice pour le patient reste supérieur aux risques engendrés. Néanmoins, bien que le risque d'induire un cancer ne puisse être négligé, le risque majeur pour le patient, dans la mesure où le processus de justification est respecté, est la non-détection d'une lésion pathologique.

Le but de ce travail était de proposer une stratégie pour optimiser l'exposition des patients tout en maintenant la précision du diagnostic en utilisant une méthodologie pertinente dans un contexte clinique.

Dans ce contexte, l'analyse objective de qualité d'image devrait tenir compte des quatre éléments suivants: (1) elle devrait être liée à une tâche; (2) les propriétés des signaux et des milieux doivent être définies en fonction de leurs propriétés statistiques; (3) l'observateur doit être spécifié et (4) une figure de mérite doit être définie. Ainsi, les modèles d'observateurs, outils mathématiques utilisés comme substitut aux observateurs humains, sont performant pour estimer objectivement la qualité d'image et répondre à une tâche diagnostique précise. Les modèles d'observateurs peuvent en effet effectuer une tâche (par exemple, détection de lésion) pour un type d'image et un signal (par exemple, un fond uniforme mais bruité) et permettre une estimation quantitative de la performance (par exemple, l'aire sous la courbe ROC).

En outre, l'avantage des modèles d'observateurs est qu'ils sont économiques, en terme de temps et d'argent, et sont également consistant dans leurs réponses contrairement aux observateurs humains.

Ce travail montre que l'utilisation d'une approche axée sur la tâche clinique pour comparer les unités de TDM et les protocoles cliniques en terme de qualité d'image et d'exposition des patients devient réalisable grâce aux modèles d'observateurs. Une telle approche donne l'opportunité d'estimer le potentiel de réduction de dose offert par les derniers développements technologiques. Elle permet aux physiciens médicaux de convertir les informations cliniquement pertinentes définies par les radiologues en des critères de qualité d'image.

Abstract

The technological evolutions of diagnostic X-ray imaging modalities enable to radiologists improve diagnosis quality and patient care. In this context, the number of X-ray examinations like conventional radiography, fluoroscopy or computed tomography (CT), is increasingly used in patient care. The risk associated with the use of ionizing radiation in medical imaging is the risk of inducing cancer, a risk which is by the Linear No-Threshold model traditionally developed for patient radiation protection. In addition, CT imaging contributes to roughly 70 % of the total annual effective dose delivered by X-ray imaging to the population. Because of this, many efforts have been made to decrease patient exposure to ensure that the risk benefit balance clearly lies on the benefit side. Nevertheless, while the risk of inducing cancer cannot be neglected, the major risk for the patient, if the justification process is respected, was the non-detection of a pathological lesion.

The goal of this work was to propose a strategy to optimise patient exposure while maintaining diagnostic accuracy using a task-based methodology that is pertinent in a clinical context when dealing with CT imaging.

In this context, objective image quality should be developed and should take into account the following four elements: (1) It should be linked to a task; (2) the properties of signals and backgrounds have to be defined in accordance with their statistical properties; (3) the observer should be specified and (4) a figure of merit should be precisely defined and quantified. In this sense, model observers, which are mathematical tools potentially used as a surrogate for human observers are well suited to objectively estimate image quality at the diagnostic accuracy level. They can indeed perform a task (e.g. lesion detection) for a given type of image and signal (e.g. noisy uniform background) and allow a quantitative performance estimation using for example the area under the receiver operating characteristic curve. In addition, the advantage of model observers is that they are economical, both in terms of time and money and they are consistent unlike the human observers.

This work shows that using a task-based approach to benchmark CT units and clinical protocols in terms of image quality and patient exposure becomes feasible with model observers. Such an approach may be useful for adequately and quantitatively comparing clinically relevant image quality and to estimate the potential for further dose reductions offered by the latest technological developments.

The methodology developed during this PhD thesis enables medical physicists to convert clinically relevant information defined by radiologists into task-based image quality criteria.

Abbreviations

ASIR	adaptive statistical iterative reconstruction
ASiR-V	adaptive statistical iterative reconstruction-V
AUC	area under the ROC curve
ATCM	automatic tube current modulation
BSS	basic safety standard
$CTDI_c$	centre
CHO	channelized Hotelling model observer
CT	computed tomography
$CTDI_{vol}$	computed tomography dose index
CSF	contrast sensitivity function
CNR	contrast to noise ratio
K	covariance matrix
d_a or d'	detectability index
DRL	diagnostic reference levels
D-DoG	difference of Gaussian
DLP	dose length product
E	effective dose
EROC	estimation ROC studies
FN	false negative
FP	false-positive
FPF	false-positive fraction
FOM	figures of merit
FBP	filtered back projection
FROC	free-response ROC studies
HO	Hotelling observer
IAEA	international atomic energy agency
ICRP	international commission on radiological protection
IEC	international electrotechnical commission
IR	iterative reconstruction
LSF	line spread function
LNT	linear no-threshold model
LROC	localisation ROC studies
MO	model observers
MTF	modulation transfer function
M-AFC	multi-alternative forced choice
NPV	negative predictive value
NPS	noise power spectrum
NPW	non-prewhitening matched filter
NPWE	non-prewhitening model observer with an eye filter
PC	percent correct
$CTDI_p$	periphery of the CTDI
PPV	positive predictive value
PW	pre-whitening
PDF	probability density functions
ROC	receiver operating characteristic
ROI	region of interest
RP	resolution preference
SNR	signal-to-noise ratio
SSDE	size specific dose estimator
TTF	task transfer function
TN	true negative
TP	true positive
TPF	true-positive fraction
$CTDI_w$	weighted CTDI

List of figures and tables

1 Introduction

The technological evolution of diagnostic X-ray imaging modalities has enabled radiologists to have access to increasingly efficient systems, improving the quality of diagnosis and patient care. In this context, diagnostic X-ray imaging modalities like conventional radiography, fluoroscopy, nuclear medicine or computed tomography (CT), are increasingly used for patient care. Because of this, however, diagnostic X-ray imaging nowadays contributes from 25 to 50 % of the total annual effective dose of the population of western countries [1]. In 2007, in Switzerland or in France, the average effective dose per inhabitant due to X-ray imaging was about 1.2 mSv but increased to 1.4 mSv in 2013 [2,3] with an expected further increase to come. In comparison, in Germany, the average dose per inhabitant due to x-ray imaging was about 1.7 mSv in 2010 with 1.15 mSv due to CT imaging [4]. It is worth mentioning that the contribution of the collective effective dose of the different modalities mentioned above is quite uneven. In Switzerland in 2013, for example, about 1.0 mSv was due to CT examinations. That represented 70% of the collective effective dose delivered to the population whereas it represented only 10 % of the number of examinations[2-6] (Figure 1).

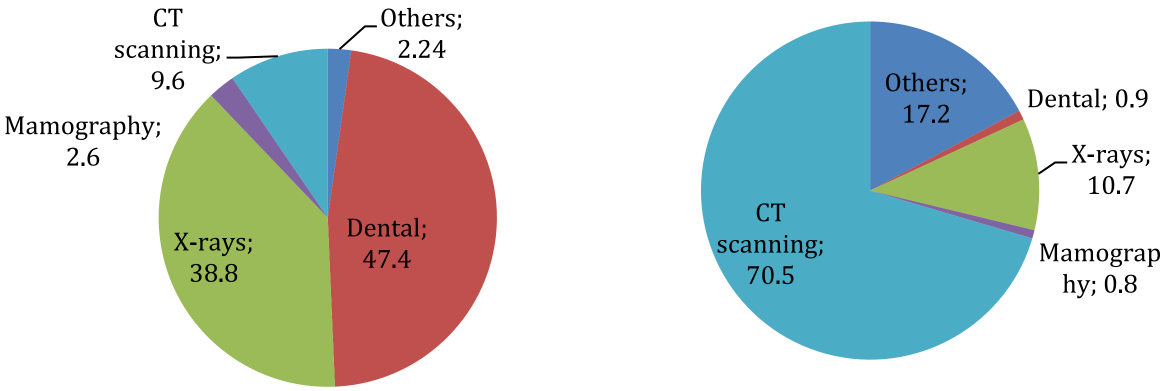

Figure 1: Summary of frequency and dose distribution

To control this trend, two radiation protection principles must be reinforced: the justification of the examination and the optimisation of the radiological procedure. Justification means that the examination must be both medically indicated and useful. While this work focuses entirely on the second principle—optimisation—it is important to note that a more rigorous justification would be an efficient way to improve the radiation protection of the population in the field of medical imaging. Justification should not only address the usefulness of the examination, but should also provide some information concerning the radiological information required to answer the clinical question, including indications for the image quality level required, which would in turn then facilitate the optimisation process. The new Swiss Ordinance on radiation protection is now more precise on the justification aspect by applying the Basic Safety Standard (BSS) published in 2014 by the International Atomic Energy Agency (IAEA).

For the second principle, the optimization process should ensure that a CT unit is as efficient as possible to convert the radiation received by the detectors into valuable image information. To achieve this, it is essential that the acquisition protocols are optimised in order to find the best trade-off between image quality and patient exposure, whatever the patient morphology.

The major risk associated with the use of ionizing radiations in CT imaging is the risk of inducing cancer. This risk is nevertheless still debated (especially for low dose levels) because the uncertainties are too high to clearly demonstrate a link in medical imaging between dose and cancer induction [7-12]. However, recently published results show that, at the cellular level, radiation effects were detected following CT examinations [13,14]. In such a context, the precautionary principle must be applied and the Linear No-Threshold model (LNT) is the standard used in the area of patient radiation protection [15] as in the field of workers exposed to ionizing radiations. Many efforts have been made by manufacturers, medical physicists, and radiologists to optimise clinical protocols in order to ensure that the risk benefit balance clearly lies on the benefit side (Figure 2). Nevertheless, while the risk of inducing cancer cannot be neglected, the major risk for the patient, if the justification process is respected, is the non-detection of a pathological lesion. Accordingly, it is important to be sure that dose reductions do not impair the diagnostic information required by the radiologist.

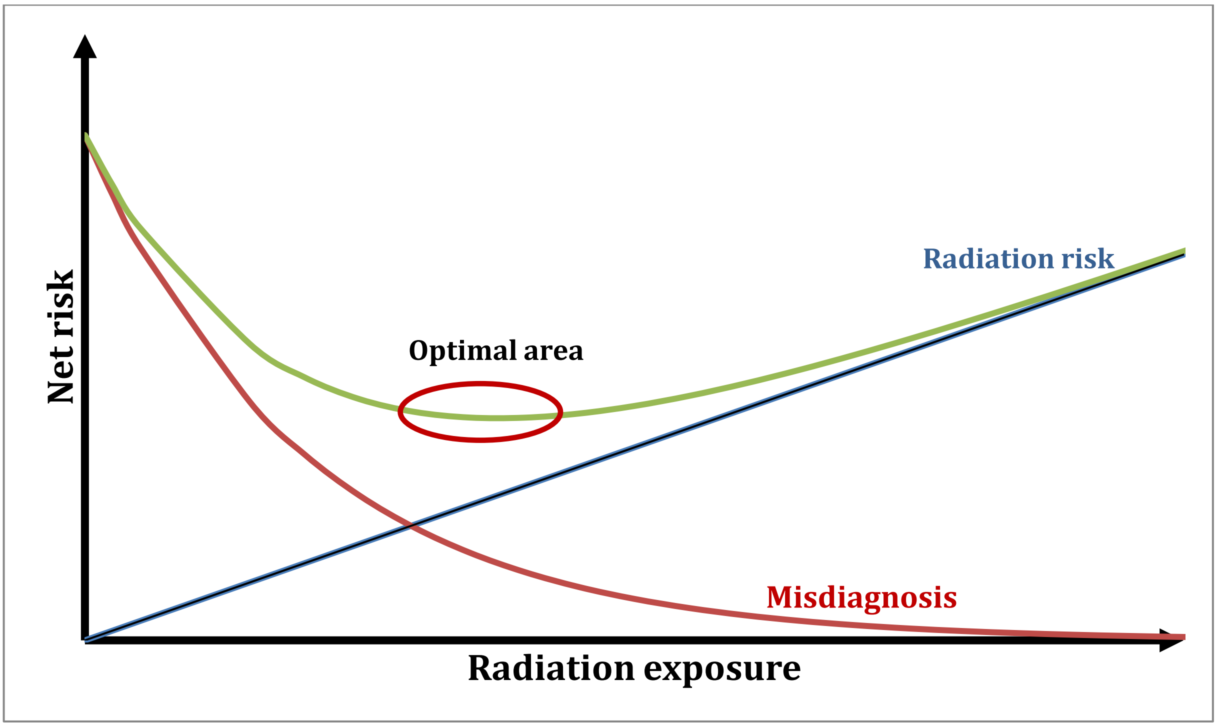

Figure 2: Trade-off between radiation risk and misdiagnosis

In response to the population's increased exposure through medical imaging, the supervisory authorities have strongly advocated for dose reductions. However, while still keeping this in mind, instead of considering the dose aspects, medical and imaging teams should first ensure that the necessary diagnostic information is contained in the images while trying to keep patient exposure as low as possible.

Image quality assessment in medicine can be complex and one way to tackle the problem is to take the approach proposed by Fryback and Thornbury [16] which proposes a "Hierarchical Model of Efficacy": from the pure technical properties of image quality (such as Signal-to-Noise Ratio (SNR), Modulation Transfer Function (MTF), Noise Power Spectrum (NPS), ...) to the impact of image information on the therapy, the patient well-being or even the societal efficacy (Table 1) [17,18].

Level	Designation
1	Technical efficacy
2	Diagnostic accuracy
3	Diagnostic thinking
4	Therapeutic efficacy
5	Outcome efficacy
6	Societal efficacy

Table 1: Six-level of Hierarchical Model of Efficacy for image quality assessment

Until now image quality assessments made by medical physicists addressed only the first level of the scale proposed by Fryback et al. [16]. One way to improve the situation would be to work at the second level of the scale using image quality criteria that are linked to a diagnostic task (task-based image quality criteria). Figure 3 from J. B. Solomon's PhD thesis shows the importance of assessing image quality in relation to a diagnostic task [19]. In the top row, the two images have different noise levels, but if the image quality is evaluated with a task-based criterion (detection of a calcified structure), the outcome will be similar. However for another task, such as the detection of a focal liver lesion (bottom row), the difference in image noise levels might lead to a different outcome.

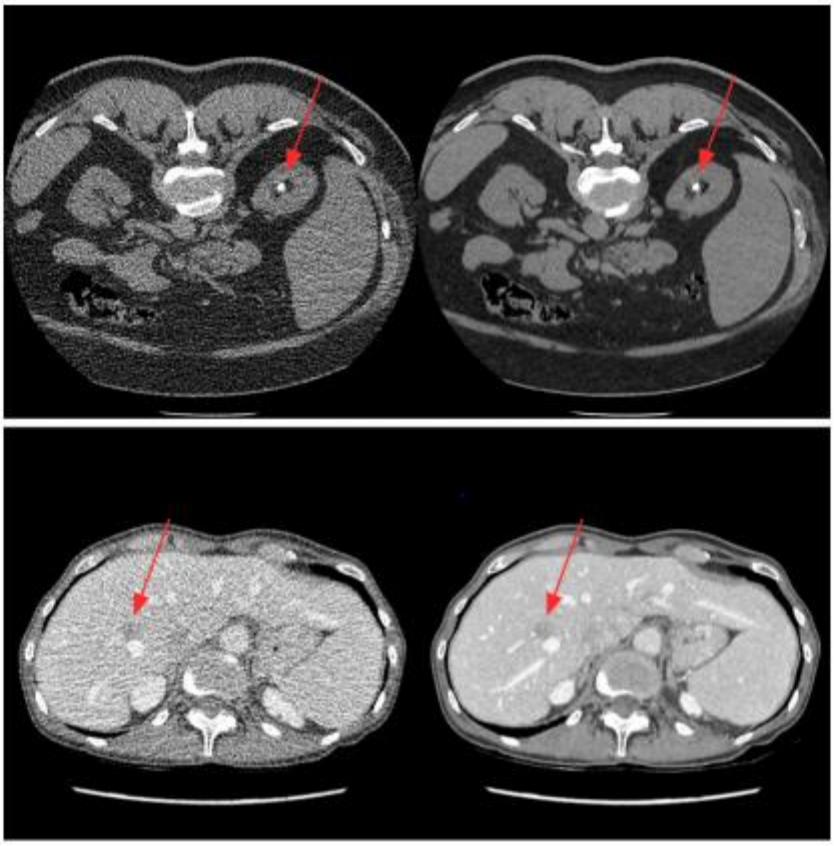

Figure 3: The top row shows two images of the same patient who underwent a CT exam due to suspected kidney stones. The bottom row shows two images in which a liver lesion is nearly rendered by the noise in the left image (from Solomon's PhD thesis)

Most of the time, clinical image quality is subjectively assessed and the overall perceived aspect of the image is of prime importance. With the filtered back projection (FBP) reconstruction technique, dose reduction is associated with an increase of image noise that leads radiologists to potential difficulties of detecting low contrast lesions. On the contrary, with iterative reconstruction (IR) techniques, a dose reduction is not systematically associated with high noise content in the images. Moreover, it is possible to get reasonably "good looking images" at a low dose level even if their diagnostic information content is quite low. These images can be adequate for some diagnoses but might lead to misdiagnoses for other indications. With the use of IR, the relationship between the object and its signal obtained from the imaging system is no more linear. This means that the traditional metrics used to objectively assess image quality must be adapted or changed. With IR, the production of reasonably "good looking images" prevents a reliable subjective assessment of image quality by radiologists. A way to address this challenge is by using clinically relevant task-based image quality criteria.

As mentioned previously, the first step of the optimisation process is to ensure that a maximum of X-rays produced by the imaging unit is converted into information. Previously, several figures of merit (FOM) were used to characterise the performances of a CT unit from an image quality point of view. FOM such as the "Q-value" (introduced by ImPACT in the UK) that combines a set of image quality and dose parameters already made it possible to evaluate and compare the performance of CT units. While this approach was quite useful during the development of CT technology, where performances between different units could vary drastically, it appears now that the sensitiveness of those methods are quite limited. Until recently, image quality was almost exclusively estimated through technical properties (SNR, MTF, NPS ...) corresponding to the 1st level of the Hierarchical Model of Efficacy. This estimation was correct because it was assumed that a good set of basic image parameters always led to good diagnostic accuracy.

The introduction and the development of IR created a new challenge in the field of image quality assessment. Limiting the CT characterisation using metrics remaining in the first level of the Hierarchical Model of Efficacy scale is not enough. In this case the image quality must be assessing at the minimum at the second level of the Hierarchical Model of Efficacy scale to ensure an adequacy between image quality and clinical needs. According to Barrett and Myers [17], objective image quality at the diagnostic accuracy level, 2nd level of Hierarchical Model of Efficacy, should take into account the four following factors: (1) It should be linked to a task; (2) the properties of signals and backgrounds have to be defined in accordance with their statistical properties; (3) the observer should be specified and (4) a FOM should be precisely defined and quantified. In this context model observers (MO), are intrinsically well suited to estimate image quality at the diagnostic accuracy level. In fact, they can perform a task (e.g. detection) for a given type of image and signal (e.g. noisy uniform background) and estimate the performance in quantitative terms, like the area under the Receiver Operating Characteristic (ROC) curve (AUC). Using similar metrics for patient dose surrogates as for the Q-value, there is a way to characterise the efficiency of CT when dealing with specific tasks. This methodology could also be taken into account when dealing with the optimisation of clinical protocols. Finally, concerning patient exposures, the volume Computed Tomography Dose Index ($CTDI_{vol}$), as well as the Dose Length Product (DLP) are used as a surrogate for patient exposure. The medical physicist has an active role in the process of dose management and quality assurance. The next challenge is to increase the participation of the medical physicists in the image quality optimisation process; to do that it is important to define relevant imaging tasks that could be used as a surrogate for image quality (Figure 4).

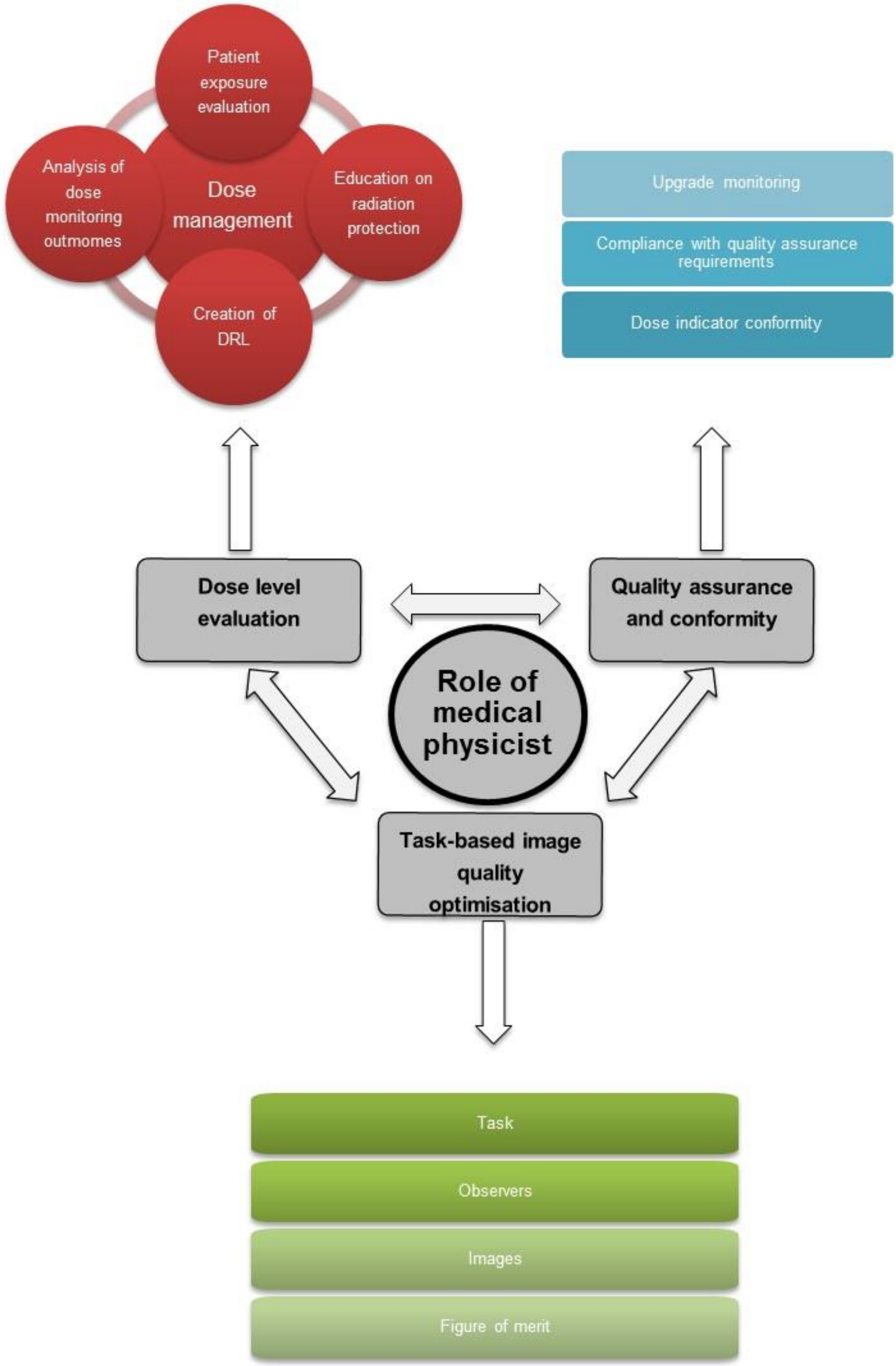

Figure 4: The role of the medical physicist in a radiology department

2 Goal of the PhD thesis

The goal of this PhD thesis was to propose a strategy to optimise patient exposure using image quality criteria that make sense in a clinical context when dealing with CT imaging. The first part of this work was devoted to defining the impact of technological developments on image quality in the field of CT imaging using a task-based approach. The second part was devoted to providing tools to measure the diagnostic accuracy of the clinical protocols. Finally, this work proposes a way to link radiologists' needs and medical physicists' tasks, trying to convert the clinically relevant information into simple task-based image quality criteria. The different methods used in this PhD thesis to achieve the different aims cited here are described in detail in the following parts.

3 Radiation dose estimation

In CT imaging, dose indicators were introduced to characterise the patient exposure and dosimetric quantities were introduced to estimate the radiological risk.

3.1 Computed Tomography Dose Index

To initiate an optimization process the first step is to provide some dose and risk indicators to the users. Many efforts have been made to better estimate the risk part of CT examinations by introducing standardised ways to quote patient exposure, for example CTDI and DLP concepts. CTDI is defined as dose profile integrated over 100 mm obtained by one rotation of the tube in axial acquisition, divided by the collimation width (see equation below). This index is used as a surrogate of the absorbed dose in the patient per scanned length unit. The scattered radiation largely overflows from the collimation and is an important part of the dose to the patient, so the CTDI must take this dose contribution into account. This is why the dose profile is integrated over 100 mm, well beyond the collimation width [20].

$$CTDI = \frac{1}{L_c}\int_{-50}^{+50} D(z)dz$$

CTDI: computed tomography dose index [mGy]
D(z): Absorbed dose profile along the longitudinal axis [mGy]
L_c: Collimation width [mm]

To evaluate the average dose in the slice, the weighted CTDI ($CTDI_w$) is defined as a weighted sum of the CTDI in the centre ($CTDI_c$) and periphery of the CTDI ($CTDI_p$) phantom, as shown in the following equation:

$$CTDI_w = \frac{1}{3}CTDI_c + \frac{2}{3}CTDI_p$$

At the end, CT manufacturers report the $CTDI_{vol}$ on the scanner control console for each examination as a requirement by the International Electrotechnical Commission (IEC) CT safety standard [21]. The $CTDI_{vol}$ is the $CTDI_w$ normalised by the pitch.

$$CTDI_{vol} = \frac{CTDI_w}{pitch}$$

To improve the scope of CT dose estimation, an adaptation of the definition of the $CTDI_{vol}$ has been proposed to enable a qualification of CT units having wider nominal beam widths [22]. The $CTDI_{vol}$ is used as a dose indicator for patient exposure but it is particularly important to note that it quantifies the dose in a simple and homogenous phantom, and is not the actual dose delivered to the patients.

3.2 Dose Length product

DLP quantifies the total dose absorbed on the explored length, but cannot be used to assess the stochastic risk associated to an examination. The total dose delivered to a patient during a CT examination depends on the scanned length and is estimated from the $CTDI_{vol}$ with the DLP metric.

$$DLP = CTDI_{vol} \cdot L$$

DLP: Dose Length Product [mGy · cm]
CTDIvol: [mGy]
L: scan length [cm]

3.3 Effective Dose

Nowadays it is suggested that the most appropriate quantity for estimating the risk due to diagnostic imaging procedures is the radiation dose to individual organs [15]. The DLP quantifies the absorbed dose in the irradiated volume, the assessment of the stochastic risk requires taking into account the radio-sensitivity of the exposed organs by calculating the effective dose (E) [23] [24] [25]. The calculation of the effective dose is not straightforward because it requires precise knowledge of the dose absorbed by a selected organ. It also depends on the fraction of the volume of each organ exposed to in the primary field and the distribution of the scattered radiation. To simplify the estimation of the effective dose or even organ dose from simpler metrics such as $CTDI_{vol}$ and DLP, conversion factors or software, such as the one proposed by Impactscan can be used [26]. Mean conversion factors (e_{DLP}) that convert the DLP quantity into effective dose have been proposed for the most common CT scans, allowing a quick estimation of the effective dose but with a limited accuracy. The values of the e_{DLP} must also be adjusted according to age and morphology of the patient, even if effective dose is defined by the International Commission on Radiological Protection (ICRP) for a reference adult or for child of 0, 1, 5 and 10 years old [15]. A set of e_{DLP} values is given is Table 1 [27].

$$E = DLP \cdot e_{DLP}$$

E: Effective dose [mSv]
DLP: Dose Length Product [mGy · cm]
e_{DLP}: Conversion factor [$mSv \cdot mGy^{-1} \cdot cm^{-1}$]

e_{DLP} ($mSv \cdot mGy^{-1} \cdot cm^{-1}$)					
Region of body	0 year old	1 year old	5 year old	10 year old	Adult
Head and neck	0.0130	0.0085	0.0057	0.0042	0.0031
Head	0.0110	0.0067	0.0040	0.0032	0.0021
Neck	0.0170	0.0120	0.0110	0.0079	0.0059
Chest	0.0390	0.0260	0.0180	0.0130	0.0140
Abdomen	0.0490	0.0300	0.0200	0.0150	0.0150
Trunk	0.0440	0.0280	0.0190	0.0140	0.0150

Table 2: e_{DLP} conversion factor

3.4 Size Specific Dose Estimator

When the patient's size differs significantly from the diameter of the CTDI phantom (∅ 32 cm for the CTDI abdomen phantom); the average dose delivered in a slice may be significantly different from the numerical value given by the $CTDI_{vol}$. A correction of $CTDI_{vol}$ by a factor depending on the effective diameter of the patient (geometric mean between the lateral size and the thickness of the patient) is necessary. The corrected $CTDI_{vol}$ is called Size Specific Dose Estimator (SSDE) [28]. However, the SSDE cannot be used to calculate the effective dose that is determined for a standard patient.

$$SSDE = f\sqrt{(LAT \cdot AP)} \cdot CTDI_{vol}$$

SSDE: Size Specific Dose Estimator [mGy]
f: correction factor in terms of effective diameter
LAT: patient lateral size [cm]
AP: thickness of the patient [cm]

In summary, the estimation of organ dose and effective dose for various anatomies and for standard acquisition protocols in general without tube current modulation, are well documented. Less scientific data are available when dealing with tube current modulation which is the most frequent situation in clinical routine and this constitutes a strong limitation to properly estimating the absorbed dose to the organs with the automatic tube current modulation [29]. Tube high voltage variation during the acquisition and organ base modulation require a further effort in the way organ doses are estimated if the acquisition protocols are to be optimised in a realistic way.

3.5 Diagnostic Reference Level

The diagnostic reference levels (DRL) is a concept used for optimising patient exposure. They provide a reference frame but are not dose limits. DRLs make it possible to set up different plans of action or correction when patient exposure is too high in comparison to nationally or regionally accepted DRL values. In radiology, the DRL is defined as the third quartile of the distribution of dose indicators ($CTDI_{vol}$ and DLP) for a given protocol. They are obtained by organising national surveys of the practice. The limitations of current DRLs are that they are defined per anatomical region which is insufficient when willing to optimise a protocol on the basis of the clinically relevant diagnostic information.

4 Dose reduction techniques

Technologically, CT scanners have continuously evolved in terms of improving their diagnostic accuracy. Nowadays the main effort is made on dose reduction. Automatic tube current modulation and iterative reconstructions are the two main tools used in clinical routines to decrease patient exposure.

4.1 Automatic Tube Current Modulation

Since 1994 manufacturers have proposed efficient tools such as the automatic tube current modulation (ATCM) to reduce patient exposure from 15% to 53% in comparison to constant tube current [30] [31] [32] [30] [33] [34] [35]. The current modulation can be carried out along the longitudinal axis (longitudinal modulation) and/or take into account differences in absorption during the rotation of the X-ray tube (angular modulation).

To define the way the modulation works, two paradigms are used by the different manufacturers. First, the modulation may be calculated to maintain noise levels per slice close to previously introduced target values, and is used by GE and Toshiba. Second, the modulation can be calculated to maintain a constant level of overall diagnostic quality for all patient sizes with respect to a reference image, thus allowing a higher noise level for larger patients (mainly because of the intrinsic contrast generated by inter-organ fatty tissue) or lower noise for thinner patients, a technique used by Philips and Siemens. But the ATCM is relatively sensitive to the patient position in the gantry [36]. Moreover, some units have an organ based tube current modulation to spare selected organs (such as eye lens, thyroid or breast) with the possibility to modify the tube current during the acquisition [37]. Recently, tube voltage modulation has been proposed to automatically select the tube voltage as a function of patient size and diagnostic task. For example, lower tube voltage can result in an improved radiation contrast and in the same time lead to a noticeable dose reduction in acquisitions where iodine contrast material is used [38 39].

4.2 Iterative reconstruction

Historically, standard FBP was used to reconstruct CT images in a very efficient way; nowadays IR algorithms are replacing the FBP algorithm. One of the limitations of the FBP algorithm is that an equal weight is given to all data that are collected whatever their information content. This means that when the attenuation is not constant during the rotation of the tube, some noisy projections significantly impair the image quality of FBP reconstructed images. IRs gives weight to the projection according to their information content. This enables noticeable dose reductions but introduces some change in the image's texture. There are two types of IR algorithms: First, the statistical IR acts on the statistical properties of image noise; it uses a blend of FBP images with different strength levels where images are reconstructed in the raw data domain and in the image domain to reduce image noise [40 41 42 43 44 45 46 47 48 49 50 51].

The other category of IR is a statistical model based IR; this algorithm uses a refined local image noise model that predicts the variance of the image noise in different directions in each image pixel and adjusts the space-variant regularization function correspondingly. The anisotropic noise model in each image pixel is obtained by analyzing the statistical significance of the raw data contributing to that pixel (in the raw data sinogram). It is of note that these iterative reconstructions work in general like black boxes. The solutions proposed might use some statistical properties of the data (by putting, for example, more weight on the intense rays rather than on the highly attenuated rays where the noise level is high) but in the end, all these solutions modify the image texture.

Finally, GE developed a full model based IR with the commercial name of VEO. Unlike its first iterative reconstruction Adaptive Statistical Iterative Reconstruction (ASIR), VEO is a fully iterative method that not only considers the data statistics but also the geometry of the machine itself by taking into account the voxel volumes of the scanned object, the focal spot size, the active area size of the detector; furthermore, iterations take place back and forth in the sinogram and image domains, converging gradually towards an optimised image "solution". Moreover, to enhance model precision of the CT scanner, complex mathematical formulations were determined to account for physical effects such as beam hardening, scatter and metal attenuation artefacts. Due to its complexity and specific properties, today VEO is only designed for acquisitions performed with the Discovery CT750 HD scanner (GE Healthcare, Waukesha, WI, USA) and a significant reconstruction time is still required to reconstruct CT images (over 30 to 45 minutes for a set of one hundred images) but dose reduction by factors as large as 3 to 7 might be possible without losing diagnostic information [52 53 54].

Manufacturers	Statistical IR	Statistical model based IR	Model-based IR
GE	ASIR	ASiR-V	VEO
Philips	iDose[4]	IMR	
Siemens	IRIS / SAFIRE	ADMIRE	
Toshiba	QDS+ / AIDR 3D	FIRST	

Table 3 : Classification of the different commercially iterative algorithms

IRs may indeed drastically reduce patient exposure. Their use, however, poses a severe problem terms of on the image quality assessment because these new algorithms are no longer linear and their behaviour is image content dependent.

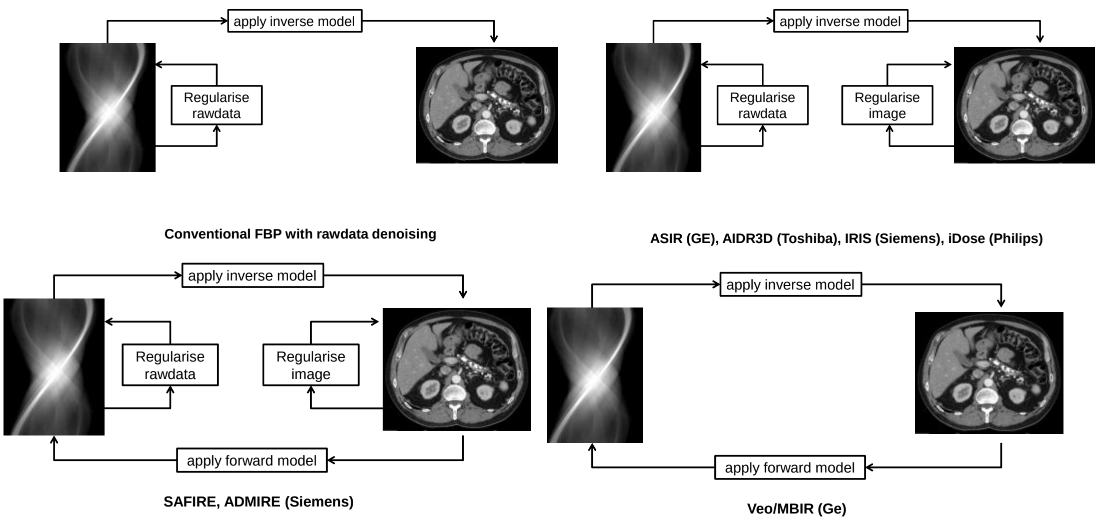

Figure 5: Schematic process of the different IR

5 Technological efficacy (1st level of Hierarchical Model of Efficacy)

The technical efficacy of diagnostic imaging concerns the physical parameters describing technical image quality in an imaging system.

5.1 Physical metrics in the image domain

In CT, different image quality metrics are used in the spatial and frequency domains [20 55]. Noise is a key parameter of image quality when dealing with the detection of low-contrast structures. Noise is voxel value fluctuation from one voxel value to another around the average voxel value in a homogeneous background. This phenomenon has two sources: the quantum noise related to the randomness of the number of photons emitted and detected per voxel, and the noise added by the electronic system (signal amplification and readout). To simply quantify the amount of noise in the image the standard deviation of voxel values (σ_x) in a homogeneous area is used.

$$\sigma_x = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{x_i} - \bar{\mathrm{x}})^2}$$

In practice, the signal corresponds to the average pixel attenuation measured in a region of interest (ROI) and the noise is computed as to the standard deviation of the pixels. From the noise and signal values, the signal to noise ratio (SNR) and contrast to noise ratio (CNR) can be obtained:

$$SNR = \frac{\bar{\mathrm{x}}}{\sigma_x} \qquad\qquad CNR = \frac{|\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2|}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

Where SNR = Signal to Noise Ratio, $\bar{\mathrm{x}}$ mean voxel value and σ_x is the standard deviation.
When dealing with the detection of small structures the spatial resolution properties of the image are of prime importance. It can be characterised by the assessment of the line spread function (LSF) in the image domain, but the use of the Fourier domain is generally preferred.

5.2 Physical metrics in the Fourier domain

Image quality can also be evaluated in the Fourier domain where the spatial resolution is materialised by the MTF. The MTF is defined as the Fourier transform of the LSF. The LSF is obtained by calculating the first derivative of the edge spread function that represents the response of the CT machines to a step in contrast of a test object. Generally the MTF is normalised by its zero-frequency value. This metric is an objective characterization of the spatial resolution and describes how well frequencies are transferred by the system.

$$MTF(u,v) \stackrel{\text{def}}{=} \frac{|FT\{LSF(x,y)\}|}{|FT\{LSF(0,0)|}$$

Where FT represents the Fourier transform and LSF (0,0) is line spread function at the zero spatial frequency.

The concept of MTF is generalised by introducing a metric called task transfer function (TTF).The spatial resolution evaluated using the TTF takes into account the effect of contrast on the reconstruction with iterative algorithm. 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 [56 57]. The image noise can be characterised by the noise power spectrum (NPS). Assuming that the noise is stationary across the image, the NPS gives a complete description of the noise by providing its amplitude over the entire range of the image's frequency. The 2D NPS is calculated on a flat image (fx, fy):

$$NPS_{2D}\left(f_x, f_y\right) = \frac{\Delta_x \Delta_y}{L_x L_y} \frac{1}{N_{ROI}} \sum_{i=1}^{N_{ROI}} |FT_{2D}\{ROI_i(x,y) - \overline{ROI_i}\}|^2$$

Where $\Delta_x \Delta_y$ are the sizes of the pixels in dimension x and y; Lx, Ly are ROIs length (in pixels); N_{ROI} the number of ROI used; $\overline{ROI_i}$ the mean value of the pixels of the ith ROI.

5.3 Combination between image quality and dose metrics

To compare the performance of various CT units it is possible to synthesise some of the parameters described above. For example ImPACT introduced a FOM, (the Q factor) that was used for many years on CT scanners [55]. The Q-factor balance dose and image quality in one FOM. It combines spatial resolution (f_{av} is the average of the 50% and 10% values of the MTF) noise (σ) and dose ($CTDI_w$). In addition it includes a parameter that takes into account the longitudinal resolution (z) of the acquisition:

$$Q_2 = \sqrt{\frac{\mathrm{f}_{\mathrm{av}}^3}{\sigma^2 z CTDI_w}}$$

This approach was quite simple and made is possible to compare completely different CT technologies (i.e. the difference between xenon and state-solid detectors). However, with more modern systems, the complexity of the scanners and their reconstruction processes create certain limitations concerning the sensitiveness of the technique. Additionally, the FOM is not task oriented. Using no task-based paradigm creates some bias when the image quality is evaluated. Noise texture and resolution can impact the detectability which is not highlighted with an image quality assessment that is not linked with a task like the CNR. As an example, Figure 6 (image from J. B. Solomon's PhD thesis [19]) shows images with equal CNR but different a detectability index.

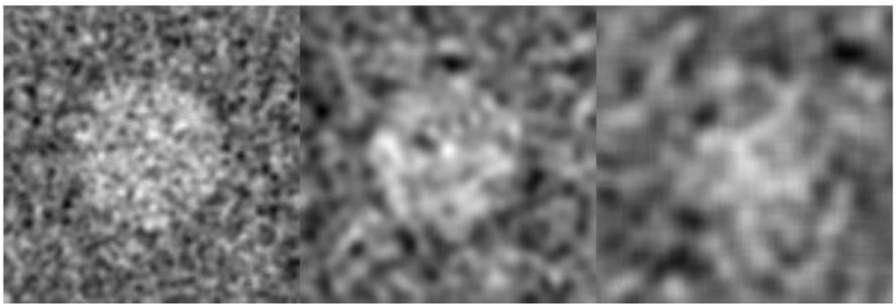

Figure 6: Equal CNR but different detectability index (image from J. B. Solomon's PhD thesis [19])

To solve the problem it is possible to use model observers, such as pre-whitening model observer (PW) or non-prewhitening model observer with an eye filter (NPWE) or human observers using objective tools linked to a task to assess the diagnostic accuracy.

6 Diagnostic accuracy efficacy in CT (2nd level of Hierarchical Model of Efficacy)

Diagnostic efficacy measures performance of the imaging for the purpose of making diagnoses and that they all require interpretation of the image by an observer.

6.1 Receiver Operating Characteristics study

Another way to objectively assess the relevant information content of images, with a discrimination strategy, is to use the ROC methodology. The goal of this method is to determine the accuracy of the diagnostic test aiming to distinguishing normal from abnormal situations based on the separation of the probability density functions (PDF) of the two corresponding classes (figure 7). The results of such a binary test can be summarised in a 2x2 table (
table 4) that contains the four possible decisions, two of them correct and two of them incorrect [58]. If the outcome is correct it can either be true positive (TP) or true negative (TN) depending on whether the prediction was abnormal or normal respectively.

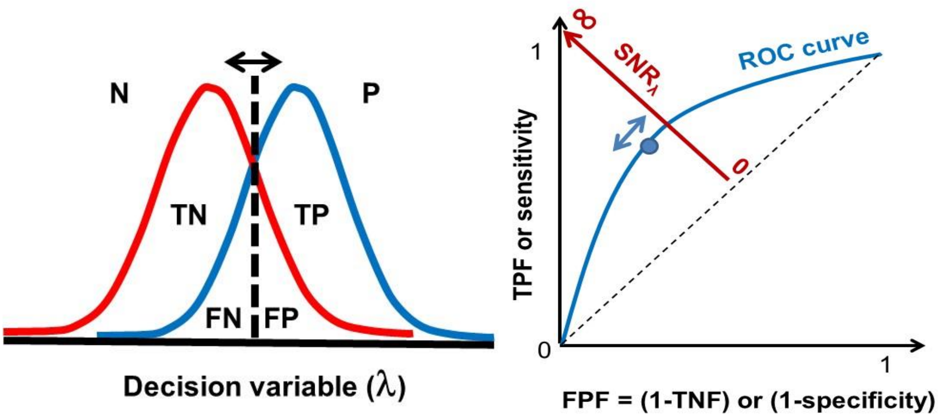

Figure 7: The distribution of the classes (negative and positive) is shown. By varying the decision threshold the number, representing the four classes will change and plotting TFP (y-axis) versus FPF (x-axis) will result in a ROC curve.

In the same way incorrect responses can be false positive (FP) (prediction was abnormal but outcome is normal) or false negative (FN) (prediction was normal but the outcome abnormal).

	Actually Abnormal	Actually Normal
Diagnosed as Abnormal	True positive (TP)	False positive (FP)
Diagnosed as Normal	False negative (FN)	True negative (TN)

Table 4: The table shows the four classes with respect to a diagnostic test.

Using these decisions outcomes, two important quantities can be defined: the sensitivity and specificity which are related to the true-positive fraction (TPF) and the false-positive fraction (FPF) using the following equation:

$$TPF = \frac{TP}{TP + FN} = Sensitivity$$

$$FPF = \frac{FP}{TN + FP} = 1 - \frac{TN}{TN + FP} = 1 - Specificity$$

These figures form the basis for the ROC curve, which is often used as a performance measure of a diagnostic test (Figure 7). The accuracy can be defined as the proportion of correct decisions out of all test subjects ($accuracy = \frac{TP+TN}{TP+TN+FP+FN}$). In addition, the positive predictive value (PPV) is the ratio of actual abnormal cases diagnosed as abnormal and the total number of cases diagnosed as abnormal ($PPV = \frac{TP}{TP+FP}$) while the negative predictive value (NPV) is defined as the ratio of actual normal cases diagnosed as normal and the total number of cases diagnosed as normal ($NPV = \frac{TN}{TN+FN}$). It should be emphasised that in contrast to TPF and FPF, the PPV, NPV and accuracy are all dependent on class prevalence. This implies that if two identical studies are performed in two different places with similar populations but with a different disease prevalence, different performances will be reported in terms of PPV, NPV and accuracy [59].

To summarise the information obtained from a ROC study the AUC is generally determined as figure of merit. The AUC varies from 0.5, where the observer does not perform better than chance to 1.0, where the observer is perfect. The detectability, d_A, related to a rating scale experiment can be derived from the AUC:

$$d_A = \sqrt{2}\Phi^{-1}(AUC)$$

where, $\Phi = \int_{-\infty}^{x} \phi(y)dy$ is the cumulative Gaussian function and $\phi = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ a Gaussian function. The d_A index varies from 0 to infinity.

Evaluating the clinical image quality using ROC theory is based on the truth. The truth can be defined in two ways: either the truth is known exactly, in that case the truth is called the ground truth, or the truth is based on various experts' decision or other pathology tests, and in this case the truth is called the gold standard. The ROC studies can be generalised to Localisation ROC studies (LROC), Estimation ROC studies (EROC) or free-response ROC studies (FROC) [60] [61] [62].

6.2 Multi-Alternative Forced Choice

In forced choice experiments, the observer has to make the 'signal present' decision between alternatives a set of offered, even if this means that he has to guess. Compared to ROC studies multi-alternative forced choice (M-AFC) experiments ("M" being the number of images that the observer has to consider to make his/her choice) are faster and easier to perform [63] but do not provide insight into the underlying distribution functions and the trade-off between sensitivity and specificity [58]. Therefore, M-AFC are sometimes referred to as a poor measure of sensitivity [64].

The natural outcome of M-AFC experiments is a percent correct (PC). For 2-AFC experiments, the PC is equal to the AUC but with human observers.

A detailed comparison and discussion about the use of ROC and M-AFC experiments as well as the optimum selection of M has been presented by Burgess [63]. Most commonly M has a value of 2 or 4 but Burgess has demonstrated that a higher value of M will result in a smaller coefficient of variance [63]. Finally, when designing M-AFC experiments care should be taken to avoid bias (e.g. the observer tends to choose left when he is unsure) [65]. An example of a trial used during a m-AFC study is shown in Figure 8.

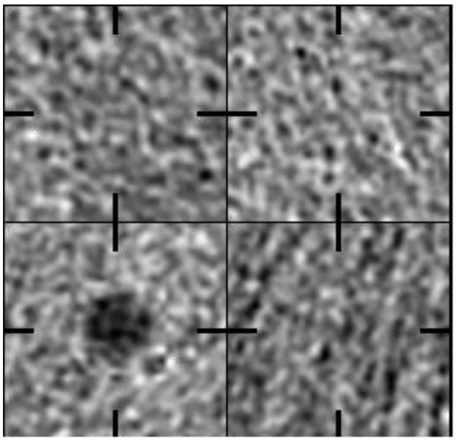

Figure 8: An example of a trial of a 4-AFC study, the signal is localised at the bottom left

Unfortunately, human observer studies such as ROC or M-AFC studies are time consuming, expensive and the inter- and intra-observer variability is often large. One way to speed up the process is to use mathematical model observers as a surrogate for human observers.

7 Model observers: A surrogate to the human observer

7.1 Task-based assessment

To evaluate the image quality in a framework of patient exposure optimization, the use of a task-based paradigm could be a way to establish a bridge between the worlds of radiologists and medical physicists. With such a paradigm four items must be defined:

The task: The task can be a classification task (i.e. detection task) or an estimation task (i.e. lesion size estimation). Often the task is linked to a single structure but there are several differences between the actual structure and its reproduction with an imaging system. First of all, a structure can be represented by a function of continuous variables, whereas the image obtained from a system is a set of discrete numbers.

The properties of signal (for example: structure to be detected) and background: The image quality assessment should take into account the physical and statistical properties of the signal and background. For example, in classification tasks (normal / abnormal), the ensemble of images that represents the hypothesis, signal present/signal absent, constitutes two populations where all the statistical variations are represented leading to the full probability density function.

The observer: To assess the image quality an observer has to be defined. This observer can be a human observer (i.e. medical physicist, radiologist ...) or a mathematical observer (i.e. model observer). Mathematical observers can be used as surrogates for human observers especially when dealing with the optimization of an imaging system which is time consuming and thus expensive (i.e. conventional ROC studies). Moreover the intra-variability with model observers is negligible, the main challenge being the choice of the right model.

The figure of merit: After deciding the task, the structure to be detected and the type of observer, it is necessary to characterise the outcome by a figure of merit and its variance that characterises the performance. The FOM can be an AUC, a PC or a detectability index (d_a or d').

7.2 Ideal observer

The Bayesian or ideal observer is a particular observer since it utilises all information available in images to maximise the performance of a given task.

7.2.1 General expression of the ideal observer

The ideal observer can be directly derived by minimizing the mean cost defined in (1) with two basic assumptions [66]. The first assumption is that a decision is deterministic. In other words, it means that for given subset Γ_i of all possible images **g**, the observer will always give the same answer D_i:

$$P\left(D_i \middle| H_j\right) = \int_{\Gamma_i} d^M g \, p\left(\mathbf{g} \middle| H_j\right) \qquad \textbf{(1)}$$

where M is the number of pixels of the image.

The second assumption is that the observer is forced to make a decision whatever the image **g** that belongs to reality H_j. Mathematically, this is translated into:

$$\int_{\Gamma_0} d^M g \, p\left(\mathbf{g} \middle| H_j\right) + \int_{\Gamma_1} d^M g \, p\left(\mathbf{g} \middle| H_j\right) = 1 \qquad \textbf{(2)}$$

where Γ_0 is the subset of image space that will lead to decision D_0 and Γ_1 is the subset of image space that will lead to decision D_1. Furthermore $\Gamma_0 + \Gamma_1$ equals the ensemble of all possible images **g**. 0 represents the index for signal absent image and 1 represents the index for signal present images.

The ideal observer can be obtained by minimizing the mean cost defined in (2). In other words, choose the subset Γ_1 that minimises the mean cost of the decision. The mean cost can be rewritten as:

$$\bar{C} = C_{00}P(H_0) + C_{01}P(H_1) + \int_{\Gamma_1} d^M g \left[(C_{10} - C_{00}) p\left(\mathbf{g} \middle| H_0\right) P(H_0) + (C_{11} - C_{01}) p\left(\mathbf{g} \middle| H_1\right) P(H_1) \right] \qquad \textbf{(3)}$$

Because the costs and the prevalence are constant, the expression of Γ_1 that minimises the cost can be obtained by only including into Γ_1 the images that produce a negative argument in the integral of (2). This leads to D_1 each time the observed image **g** is such that:

$$(C_{10} - C_{00}) p\left(\mathbf{g} \middle| H_0\right) P(H_0) + (C_{11} - C_{01}) p\left(\mathbf{g} \middle| H_1\right) P(H_1) < 0 \qquad \textbf{(4)}$$

Rearranging the terms leads to an observer response Λ that can be written as:

$$\Lambda(\mathbf{g}) = \frac{p\left(\mathbf{g} \middle| H_1\right)}{p\left(\mathbf{g} \middle| H_0\right)} \underset{D_0}{\overset{D_1}{\gtrless}} \frac{C_{10} - C_{00}}{C_{01} - C_{11}} \frac{p(H_0)}{p(H_1)} \qquad \textbf{(5)}$$

As Barret and Myers say, this inequality can be read "*decide hypothesis H_0 true whenever the greater-than sign holds; decide hypothesis H_1 when the less-than sign holds*." [17].

Thus, the ideal observer makes its decision by computing the ratio of the likelihoods of observing the given image **g** conditional to H_1 and H_0, and by comparing the ratio to a threshold (right hand term of (5)). It is an observer that utilises all information available regarding the task to maximise the performance as defined by the mean cost but does not always give the correct answer. Varying the costs changes the decision threshold and thus the optimal operating point on the ROC curve. Minimizing the probability of error would have led to the same strategy except that the threshold would have been different.

(5 was derived by minimizing the mean cost. The same strategy, but with a different threshold, would have arisen if the probability of error had been minimised or if the Neyman-Pearson criterion had been used. In other words, any of these other minimizing criteria would result to different operating points on the same ROC curve.

7.2.2 Special case of multivariate normal images

Assuming the pixel value follows a Gaussian distribution and that the covariance between all pixels is defined by the covariance matrix **K**, the probability of observing an image **g** if hypothesis H_j is true is given by:

$$p\left(\mathbf{g}\middle|H_j\right)=\frac{1}{\left(2\pi\right)^{M/2}\sqrt{\det\left(\mathbf{K}_j\right)}}e^{-\frac{1}{2}\left(\mathbf{g}-\bar{\mathbf{g}}\right)^T\mathbf{K}_j^{-1}\left(\mathbf{g}-\bar{\mathbf{g}}\right)} \tag{6}$$

where det computes the determinant of a matrix, "^{T}" is the transpose operation, $\bar{\mathbf{g}}_j$ is the mean value of **g** under hypothesis j and $\mathbf{K}_j$ is the covariance matrix of the images that belong to category j. The response of the ideal observer can be rewritten in this special case by inserting (6) into (5) and by recognizing that the logarithm function is monotonous:

$$\ln\left(\Lambda\left(\mathbf{g}\right)\right)=\lambda\left(\mathbf{g}\right)=-\tfrac{1}{2}\left(\mathbf{g}-\bar{\mathbf{g}}_1\right)^T\mathbf{K}_1^{-1}\left(\mathbf{g}-\bar{\mathbf{g}}_1\right)+\tfrac{1}{2}\left(\mathbf{g}-\bar{\mathbf{g}}_0\right)^T\mathbf{K}_0^{-1}\left(\mathbf{g}-\bar{\mathbf{g}}_0\right)\underset{D_0}{\overset{D_1}{\gtrless}}\lambda_c \tag{7}$$

If we further assume that the image noise (and therefore its covariance matrix **K**) is the same under both hypotheses and that the difference between the mean image that contains the signal and the mean image that does not contain the signal is equal to the searched signal $\mathbf{s}=\bar{\mathbf{g}}_1-\bar{\mathbf{g}}_0$, we obtain the following very compact expression for the ideal observer:

$$\lambda'\left(\mathbf{g}\right)=\mathbf{s}^T\mathbf{K}_n^{-1}\mathbf{g}\underset{D_0}{\overset{D_1}{\gtrless}}\lambda'_c \tag{8}$$

We see that in this case, the ideal observer is linear in terms of the image **g**. The strategy of this observer consists in first pre-whitening the signal template ($\mathbf{K}_n^{-\frac{1}{2}}\mathbf{s}$) and the image ($\mathbf{K}_n^{-\frac{1}{2}}\mathbf{g}$). Then the pre-whitened signal template is multiplied by the pre-whitened image in order to produce the scalar observer response. This is why this observer is usually called the PW model observer. The following expression represents the d' index obtain with the PW model in image domain and Fourier domain (The development of the d' index is given in annexe 1).

$$d'=\sqrt{\boldsymbol{s}^T\boldsymbol{K}_n{}^{-1}\boldsymbol{s}}$$

By analogy, we can transpose the d' index created in the image domain to the Fourier domain. In that case the covariance matrix is represented by the NPS and the signal by the contrast level convolves with the TTF and the input signal.

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

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.

At this stage, some precautions have to be taken concerning the effect of scatter radiation. The TTF is reduced by scatter, as well as the image contrast, and this effect should not be taken into effect twice. In our calculation, we made the choice of including the scatter effect using the measured contrasts rather than the nominal contrast. If the image noise is white (uncorrelated and of equal variance for each pixel) the ideal observer simplifies to:

$$\lambda'(\mathbf{g}) = \mathbf{s}^{\mathrm{T}}\mathbf{g}$$

which is called the non-prewhitening matched filter (NPW). This observer is sometimes used also in cases involving coloured noise, but it suffers then from the penalty of not including the noise decorrelation process in its detection strategy, and is therefore not ideal. The following expression represents the d' index obtained with the NPW model observer in image domain and Fourier domain (The development of the d' index is given in annexe 2).

$$d' = \sqrt{\frac{(\boldsymbol{s}^{\mathrm{T}}\Delta\boldsymbol{g})^2}{\boldsymbol{s}^{\mathrm{T}}\mathbf{K}_{\mathrm{n}}\mathbf{s}}} \qquad\qquad d' = \sqrt{2\pi}\,\Delta HU \frac{\int_0^{f_{Ny}} S^2(f)TTF^2(f)\, f\, df}{\sqrt{\int_0^{f_{Ny}} S^2(f)TTF^2(f)\, NPS(f) f\, df}}$$

In summary, in the case of multivariate normal images with the same noise under both hypotheses, the ideal observer is the PW. If the image noise is uncorrelated and statistically similar (white noise) for each pixel, the ideal observer is reduced to the NPW. In such a case the outcomes of PW and NPW will be the same.

7.2.3 Hotelling Observer

In Gaussian data with the same covariance matrix, the Hotelling observer (HO) is equal to the ideal observer. It is an ideal observer in the sense that it maximises the SNR. However, when data are not Gaussian, the ideal observer is usually non linear. The advantage of the HO is that only the knowledge of the first and the second order statistics (mean and variance) of the data is required to extract the maximum amount of information from the image. When the mean and covariance for the image are known, the HO's template is defined as:

$$\mathbf{w} = \boldsymbol{K}_n{}^{-1}\Delta\boldsymbol{g}$$

where K_n is the covariance matrix and Δg the mean image.
The decision variable and the detectability index computed with the HO are:

$$\lambda(\mathbf{g}) = \mathbf{w}^{\mathrm{T}}\mathbf{g} \qquad\qquad d' = \sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{s}}$$

7.3 Anthropomorphic observer

The advantage of anthropomorphic model observer is their capacity to mimic human performances. Several results show that they have a large potential to provide radiologists a way to control the image quality level of their acquisitions [67 68 69].

7.3.1 Non-prewhitening model observer with an eye filter

As opposed to the PW model, the NPW model does not include the noise decorrelation process and it can be transformed into an anthropomorphic model observer by adding an eye filter function. This filter mimics the contrast sensitivity function (CSF) of the human eye. The CSF describes the sensitivity of the human visual system as a function of the spatial frequency and can be modelled by a band pass filter with a maximum at 2 to 4 cycles per degree, falling off at low and high frequencies.

$$\lambda = \mathrm{s}^{\mathrm{T}}\mathrm{E}^{2}\mathrm{s}$$

$$d' = \frac{(\mathrm{s}^{\mathrm{T}}\mathrm{E}^{2}\mathrm{s})^{2}}{\mathrm{s}^{\mathrm{T}}\mathrm{E}^{2}\mathrm{K}_{\mathrm{n}}\mathrm{E}^{2}\mathrm{s}} \qquad d' = \frac{\sqrt{2\pi}\,\Delta HU\ \int_0^{f_{Ny}} S^2(f)TTF^2(f)\,VTF^2(f)f\,df}{\sqrt{\int_0^{f_{Ny}} S^2(f)TTF^2(f)\,NPS(f)VTF^4(f)f\,df}}$$

Where, VTF(f) is the visual transfer function of the human eye $\mathrm{VTF(f)} = \mathrm{f}^{1.8}\exp(-0.6\mathrm{f}^2)$ [70].

In reality it is quite complex to obtain the complete description of mean and variance, because the statistics are obtained from samples. Most of the time, the covariance matrix (K) is approximated by its estimation $\widehat{K}$. If the number of image samples is less than the number of pixels in each image, K will be singular (p 957 [17]). In reality it is impossible to obtain enough images and thus the inversion of the covariance matrix is not robust. Therefore it is necessary to reduce the dimensionality with some features like channels, as we will explain now.

7.4 Channelized Hotelling Observer

The channelized Hotelling model observer (CHO) was first introduced by K.J. Myers and H.H. Barrett [71 72]. The channels can be thought of as filters that selectively respond to different features, spatial or temporal frequency bands, or spatial orientations[73].

7.4.1 Channelization process

To reduce the dimensionality of the HO, the image is passed through a set of J channels; where J is significantly lower than N (N is the number of pixels in the image).

With the adopted notation, a channel is an 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 a NxJ matrix where each column is one of channel u_j.

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

The channel output v_i is obtained by the dot product between the channel u_j and the image g. With this process the dimensionality goes from N to J.

$$\mathbf{v}_\mathrm{i} = \left(\mathbf{u}_\mathrm{j}\right)^\mathrm{T} \mathbf{g}$$

The general definition of the CHO model is:

$$\mathbf{w}_{\mathrm{chan}} = \boldsymbol{K}_{v/n}{}^{-1} \Delta \boldsymbol{v}$$
$$\lambda\left(\mathrm{v}_\mathrm{i}\right) = \mathbf{w}_\mathrm{v}^\mathrm{T} \mathbf{v}_\mathrm{i}$$
$$d' = \sqrt{\mathbf{w}_\mathrm{v}^\mathrm{T} \boldsymbol{v}}$$

Where $K_{v/n}$ is the covariance matrix computed from channelized images, v is the data of the signal images seen through the channels.
With this process, the number of samples necessary to invert the covariance matrix becomes smaller. Moreover with the channelized mechanism, the model can be tuned either to obtain an ideal observer (the channels are then selected to extract all the information available); or an anthropomorphic model observer that mimics human observer performances (the channels are then selected to simulate the characteristics of the human visual system).

7.4.2 Ideal channelized Hotelling model observer

The ideal CHO model is quite adequate to, for example, benchmark CT units. In such a case, image quality can be easily assessed using a detection task, with a smooth radially symmetric signal, centrally peaked in a stationary background; the ideal template should be centred on the signal and rotationally symmetric. The Laguerre-Gauss channels have these characteristics and have been proposed by Barrett as ideal channels [74].
The Laguerre-Gauss channels, LG, are defined as:

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

where a_u is the width of the Gaussian function
With the LG polynomials given by:

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

Depending of the signal and background the LG channels must be tuned in terms of a_u (exponential weighting) and P to reach a maximum and become as efficient as possible. The number of channels depends on the complexity of the background [75].
When willing to benchmark clinical protocols with images assessed by human observer, ideal MO have the disadvantage to be poorly correlated with human performance because humans are not able to use all the information contained in the image, so the MO usually outperforms the human observer.

7.4.3 Anthropomorphic channelized Hotelling observer

These channelized models use channels that mimic the spatial selectivity behaviour of the human visual system.

7.4.3.1 Dense difference of Gaussian channels, D-DoG

When the target to detect is a structure with a spherical symmetry a good approximation of the human vision is the dense difference of Gaussian (D-DoG). The advantage of the D-DoG is that it uses fewer channels in comparison to other anthropomorphic channels, such as the Gabor channels. This is particularly important since the more channels to be used the more images need to be used. To properly estimate the covariance matrix of a Jx1 vector, a general rule of thumb admits that at least Jx10 realizations are necessary (e.g., 10 channels requires only 100 images)[76]
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: $\sigma_0 = 0.005$, α= 1.4 and Q = 1.67 [77].

7.4.3.2 Gabor channels

The Gabor channels were used especially when the target does not have a spherical symmetry or if the noise is oriented.

$$V(x,y) = \exp\left[-4\ln(2)\left(\frac{x^2+y^2}{w_s^2}\right)\right]\cos[2\pi f(x\cos\theta + y\sin\theta) + \beta]$$

Where f is the spatial frequency, θ is the orientation, w_s is the width equal to 0.56⁄f for a bandwidth of one octave, and β is the phase equal to 0.
A structure with a spherical symmetry can use Gabor channels with five orientations (in some cases the noise is anisotropic), seven frequencies, and one phase, resulting in 35 channels. Orientations are chosen with values ranging from 18 deg to 305 degrees. Spatial frequencies were chosen with values ranging from 0.5 to 5 cycles·deg^{-1} in steps spaced by a multiplicative factor of 1.4 [78].

7.4.3.3 Internal noise

As described above, some methods exist to fit the model and human performances. However in some cases, even if anthropomorphic channels are used with the CHO, the model overestimates human performance; to counteract this effect, an internal-noise component is generally added on the CHO model to match the human observer performance. Internal noise can be interpreted as the introduction in the model of variations in neural firing, intrinsic inconsistency in receptor response, and a loss of information during neural transmission in the human visual system [79] [80]. The paper "Evaluation of Channelized Hotelling Observer with Internal-Noise Model in a Train-Test Paradigm for Cardiac SPECT defect detection" by Brankov illustrates the procedure for selecting the internal-noise model, tuning its parameters, and using the selection criteria [81].
At the moment, research is still devoted to matching the human perception with the model observer's outcome. No set of standardised channels has been proposed yet [67] [82] [68] [69].

8 Achieved results

The peer reviewed papers and conference proceedings that compose the core of this thesis were developed around two milestones:

A physical approach based on phantom measurements was developed to create an efficient optimization of the use of the CT unit, especially when new technologies and image reconstruction techniques are involved.

A clinical approach based on physical metrics allowing a dialogue with radiologists was used to develop an optimal use of the options offered by the new image reconstruction techniques.

8.1 Physical approach: Developing methods to improve the characterization of clinical CT units and protocols

Classical metrics in the image domain such as the CNR and SNR or in the Fourier domain like the MTF have been widely used in the past to optimise clinical protocols. However, with the introduction of IR these metrics are no longer applicable. In that case, we have started to use new tools like model observers in clinical routines to evaluate image quality.

8.1.1 Assessment of low-contrast detectability in CT using different IR

8.1.1.1 FBP versus statistical algorithm versus full model based algorithm

Classical Fourier metrics like the TTF or MTF are well described in the literature and commonly used by medical physicists but it is easier to make a link to human observer with metrics that work in the image domain. In one of our studies we used model observers that work in the image domain to overcome these limitations [83]. A CHO model observer with D-DoG channel tuned with internal noise was used to mimic the human performance. This model was used on images obtained from an anthropomorphic abdominal phantom containing 5 and 8 mm diameter spheres with a contrast level of -10 and -20 HU (hypodense lesions). The phantom was scanned at 120 kV with $CTDI_{vol}$ equal to 5, 10, 15, 20 mGy and images were reconstructed using the FBP, ASIR 50% and model-based iterative reconstruction (MBIR) algorithms. For the same $CTDI_{vol}$ level and according to the CHO model and human observer, the MBIR algorithm provided the highest detectability indexes. The outcomes of human observers and the results of CHO were highly correlated whatever the dose levels, the signals considered and the algorithms used when some noise is added to the CHO model (see Figure 9).

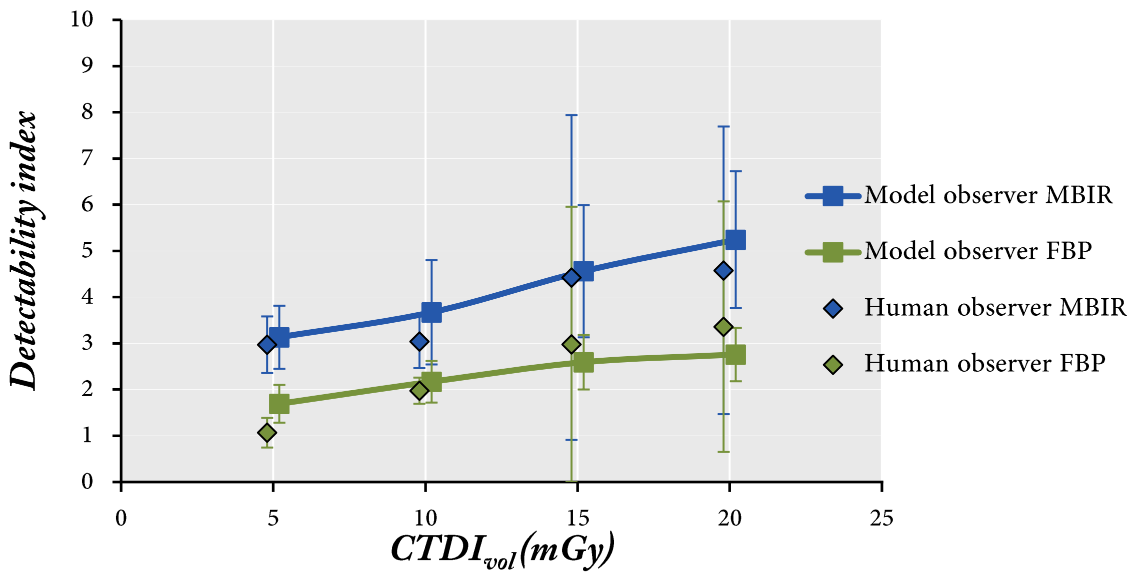

Figure 9: CHO model observer and human observers' performances with FBP and MBIR algorithms for the lesion at 8 mm and 20 HU

8.1.1.2 FBP versus hybrid model based algorithm

The same method as the one used in the study [83] mentioned previously was applied to evaluate the impact of the iterative level on image quality [84]. In this study, images were reconstructed using the iterative reconstruction method adaptive statistical iterative reconstruction-V (ASiR-V) at 0, 50 and 70 %. Internal noise ϵ was added to the decision variable λ.

$$\lambda_{noisy,i} = \lambda_i + \varepsilon_i$$

Each ϵ value is a variable obtained from a Gaussian distribution centred at zero and with a standard deviation equal to the standard deviation of the distribution of the decision variable when the signal was absent multiplied by the internal noise value α.

$$\varepsilon = \alpha \times \sigma_{bg} \times \xi$$

where $\boldsymbol{\alpha}$ is the weighting factor and ξ is a random number generated between -1 and 1, σ_{bg} is the standard deviation of the distribution of the decision variable of signal-absent.

The internal noise value α was calibrated using the signal at 6 mm 20 HU at 10 mGy. This α value was chosen to minimise the difference between the model observers and the human observer (Figure 10).

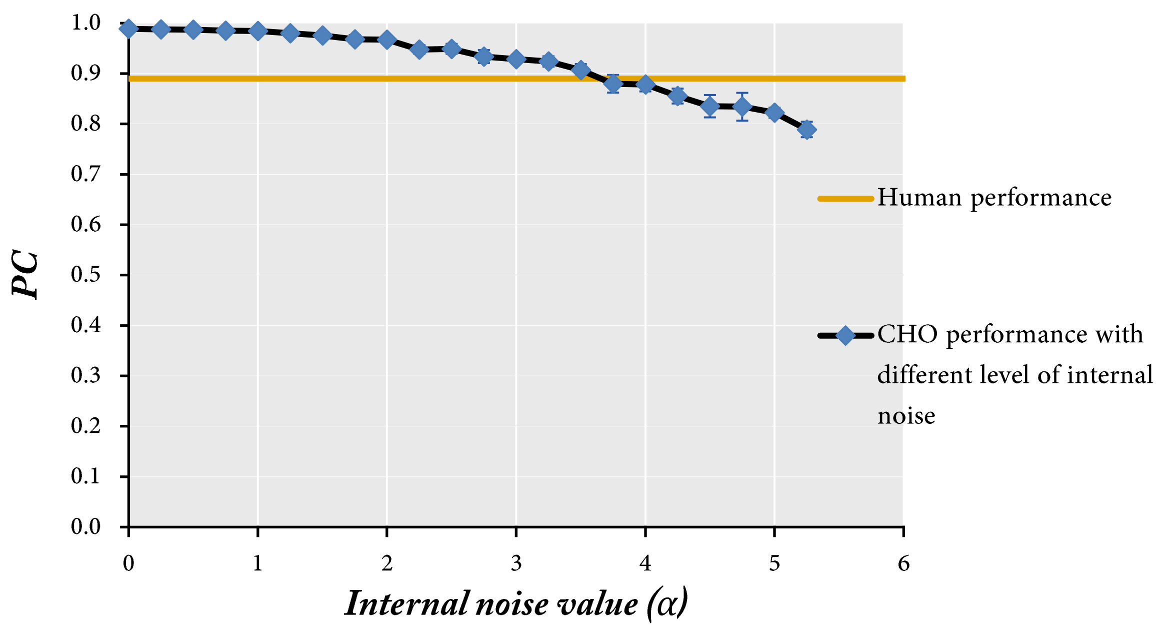

Figure 10: Calibration of the internal noise with the CHO model

The internal noise value (α=4.0) was applied to all the other categories (Figure 11).

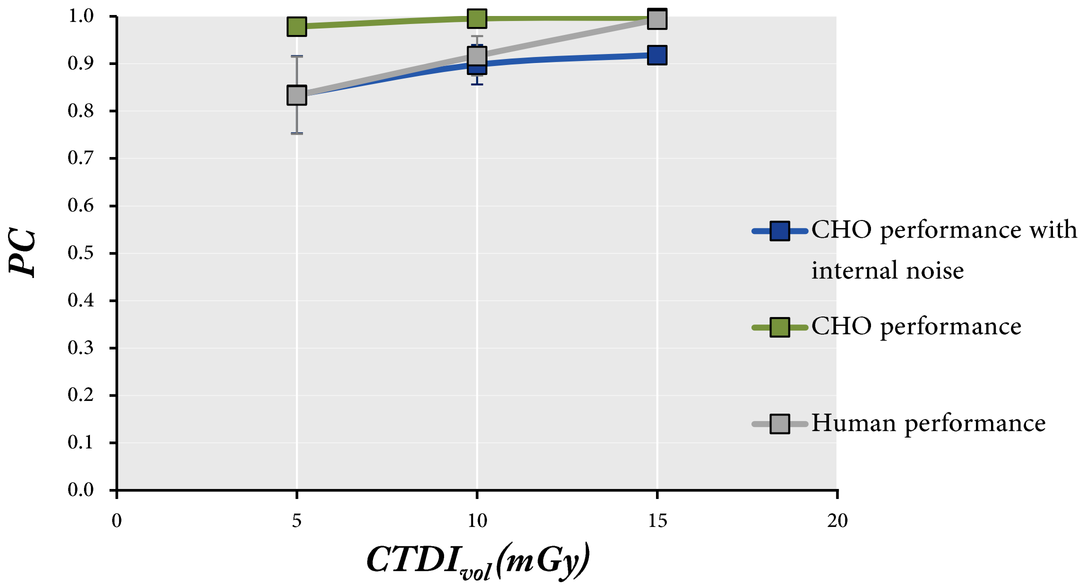

Figure 11: Human, CHO, and CHO with internal noise performance using ASiR-V at 50 % for the lesion at 6mm 20HU

An improvement in the low-contrast detectability was observed when switching from ASiR-V 0 to 50 % especially at a low dose; however, switching to ASiR-V 70% did not significantly improve the low-contrast detectability in comparison to ASiR-V 50% (Figure 12).

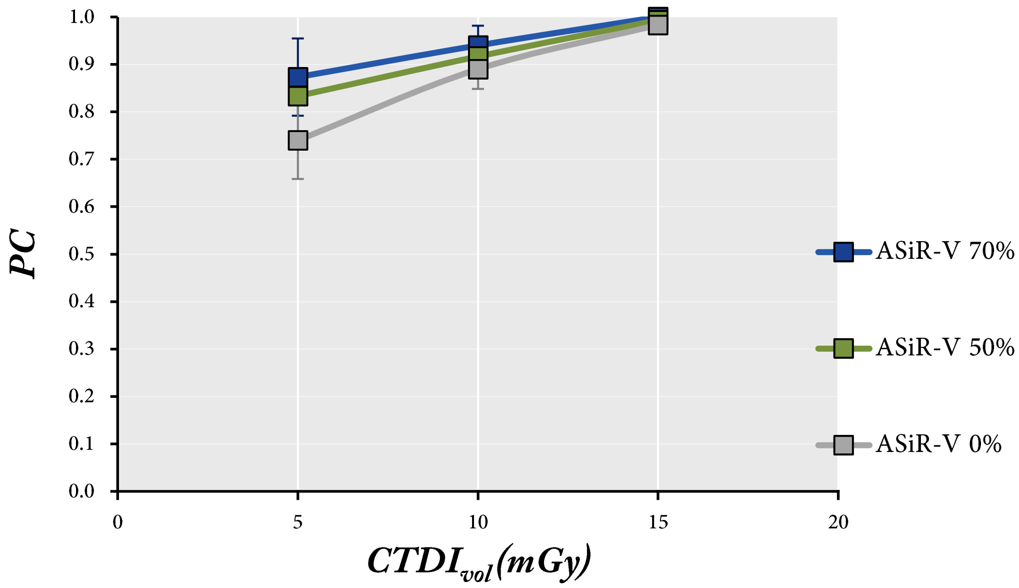

Figure 12: Human performance for the three different level at 0, 50 and 70%

8.1.2 Benchmarking of CT units

8.1.2.1 Objective comparison of high-contrast spatial resolution and low-contrast detectability on multiple CT scanners

In this study we objectively compared 8 CT scanner performances using different ideal model observers (Table 5)[85].

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

Table 5: List of CTs involved in this study

In this context using three clinically relevant protocols, the image quality was assessed using a PW model observer and a CHO model observer with Laguerre Gauss channel.

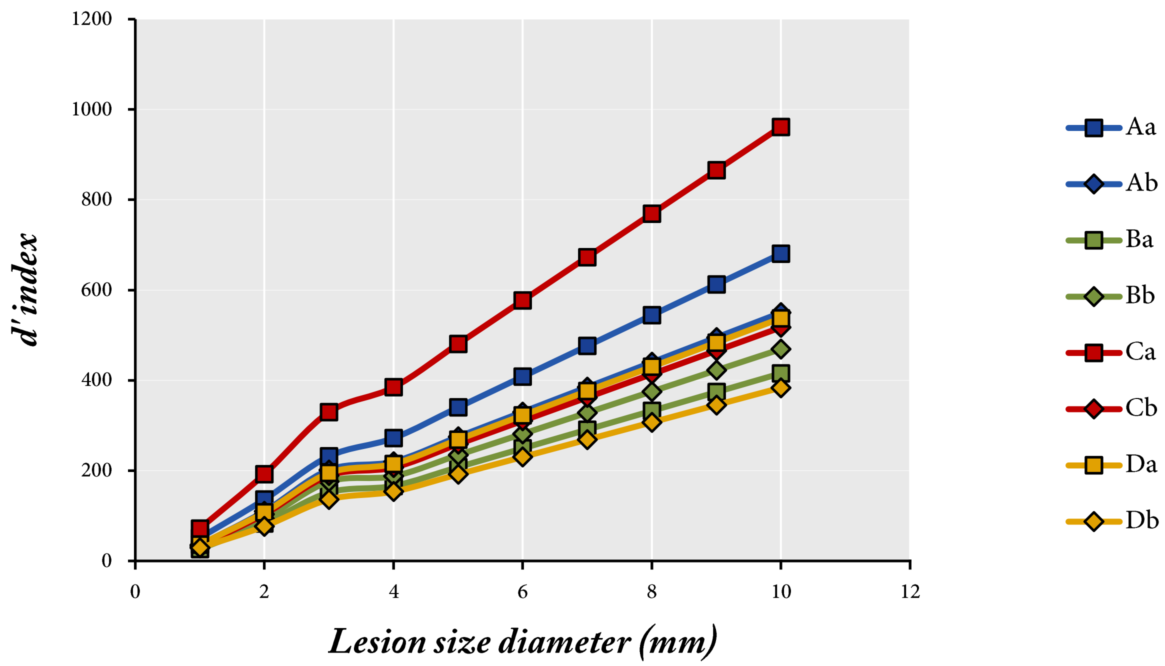

Figure 13: d' index for head protocol and contrast between PTFE and water with different lesion sizes

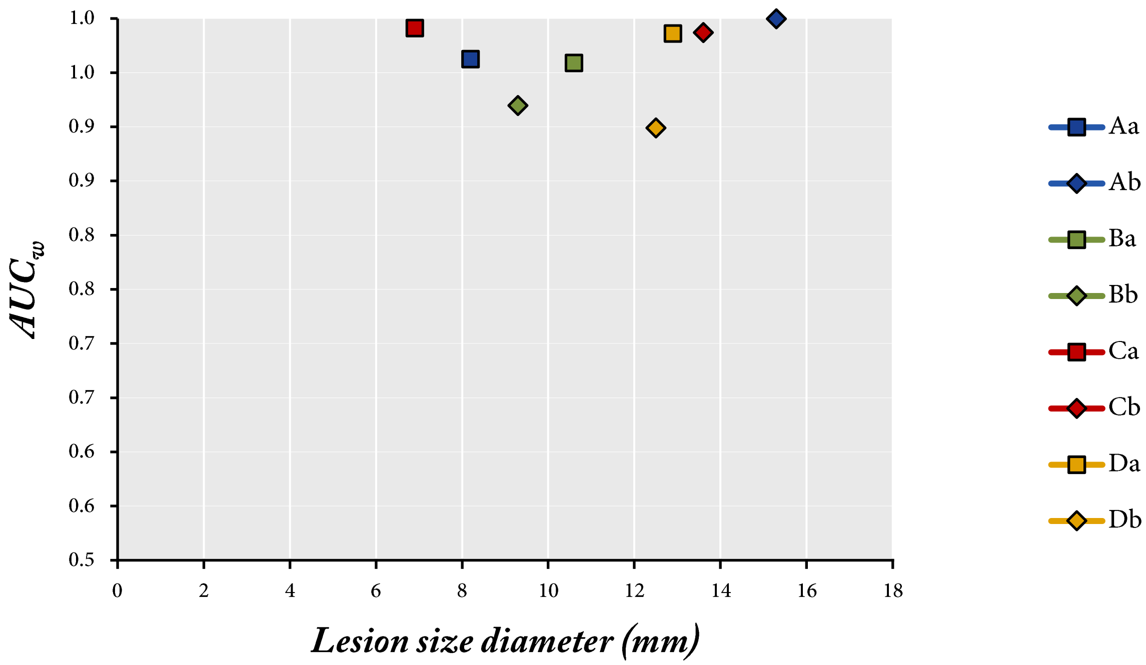

Figure 14: AUC_w for the abdomen protocol with the medium abdominal phantom

Compared with older generation CT scanners, three newer systems were found to have (Aa, Ca, Da) significant improvements in high-contrast detectability over that of their predecessors (Ab, Cb, Db). However a fourth, newer system (Ba) had a lower performance than the older CT (Bb). This study shows that MO can objectively benchmark CT scanners using a task-based image quality method, thus helping to estimate the potential for further dose reductions offered by the newer systems.

8.1.3 Benchmarking of abdominal CT protocols

8.1.3.1 Benchmarking of abdominal CT protocols using only one phantom size

Like benchmarking CT machines, a similar approach can be taken to benchmark clinical protocols [86]. In this study we used the image acquisition protocol of the portal venous phase of a multiphase abdominal protocol and we assessed the low-contrast detectability on 56 CT units, using an anthropomorphic CHO model observer on an anthropomorphic abdomen phantom. Since the clinical images are evaluated by a radiologist, it is important to use a model that mimics human observer performances. Since the spread in slice thicknesses and doses involved in the local protocols was large, an alternative metric, called 'volumetric dose', was created. The volumetric dose is defined as the product of the $CTDI_{vol}$ and the slice thickness.

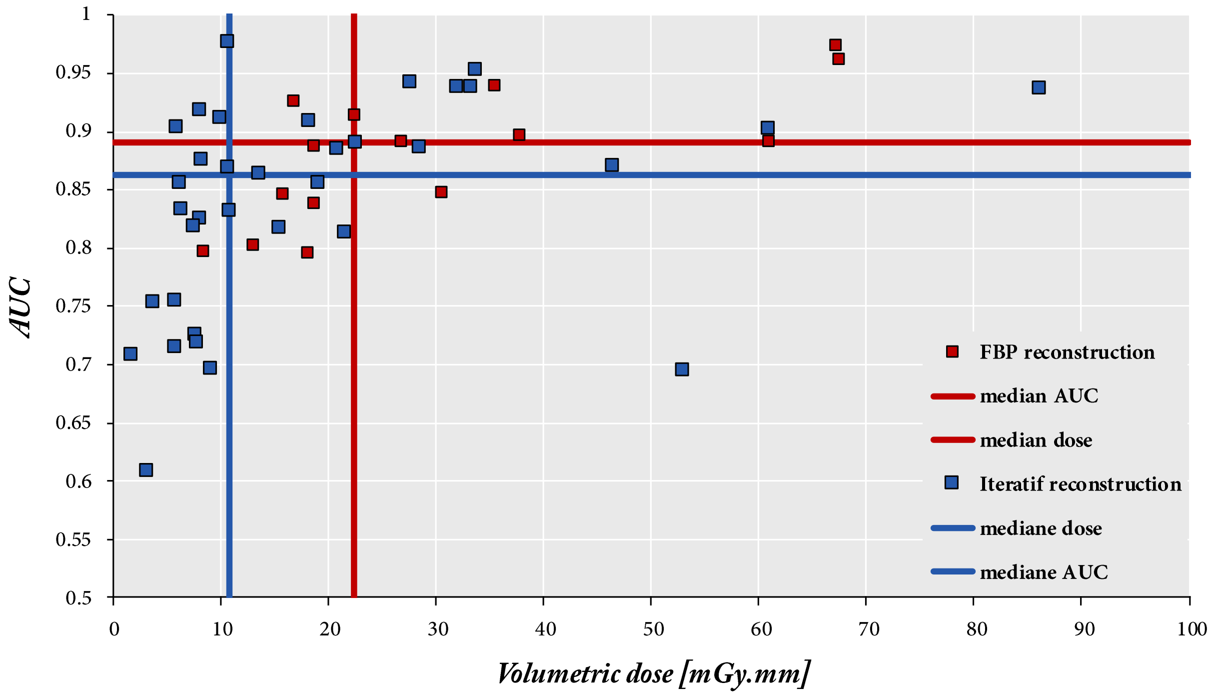

Figure 15: Results of a comparison of image quality for 5 mm/20 HU as a function of the volumetric dose.

We observed that the use of iterative reconstruction enabled a significant volumetric dose reduction (almost a factor of two) associated, however, with a slight reduction in low-contrast detectability (Figure 15). However, the main limitation of this study was the use of the volumetric dose parameter. Since partial volume effects could be very different from one protocol to another, using a volumetric dose metric was not adapted; indeed it is counter intuitive that a high volumetric dose provides a poor AUC, but this may be explained by partial volume effects, e.g. with slice thicknesses of 5 mm for the 5 or 8 mm diameter spheres.

8.1.3.2 Generalization of benchmarking of abdominal CT protocols

The characterization of clinical protocols with only one phantom size is insufficient. A good image quality for a specific patient size does not necessary mean that the image quality will be acceptable for another patient size [87]. In this study, we used three phantom sizes and investigated the practices of 68 centers. The correlation of the AUC values obtained for the different phantom sizes varied from 0.325 between the size S and L to 0.58 between the size M and L that confirm that a large variability exists when dealing with the setting of morphologically adapted protocols.

In this study, the median dose used for acquisitions was equal to 5.8 mGy, 10.5 mGy and 16.3 mGy, respectively for the small, medium and large phantoms. The median AUC obtained from acquisitions was equal to 0.96, 0.90 and 0.83, respectively for the small, medium and large phantoms. Figure 16 shows the results obtained with the medium phantom. It is interesting to note that the indication of the dose indicator with an image quality indicator facilitated discussions with radiologists proving that such an approach improved the communication between the two specialities.

Finally, our study shows that a standardization initiative could be launched to ensure comparable diagnostic information for a well-defined clinical question. We thus propose that the starting point of the optimization process be the clinical image quality levels rather than patient exposure. However, it is important to work in collaboration with radiologists, before the optimisation process, to define, the critical target to be detected and at which AUC level.

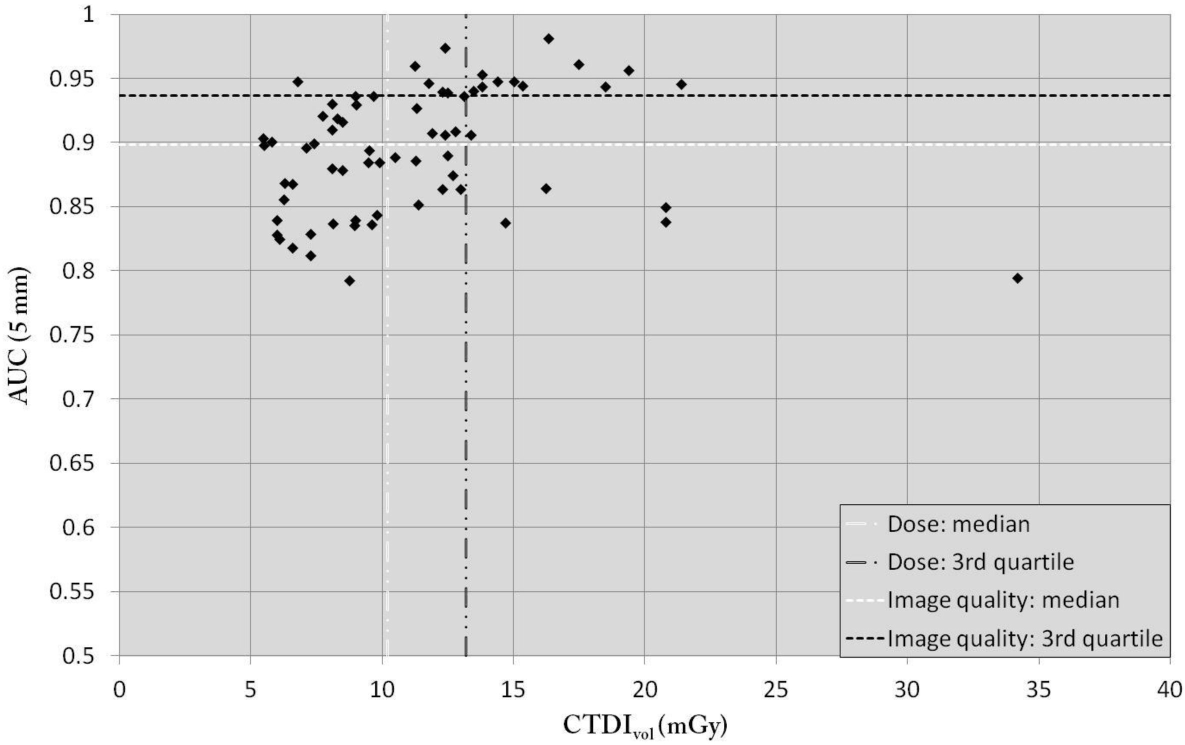

Figure 16: AUC obtained with the 5 mm lesion size as a function of $CTDI_{vol}$ for the medium size phantom.

8.2 Clinical approach: Applying methods to improve the use of IR in clinical routine

8.2.1 Optimization of IR levels for clinical thorax acquisitions

The purpose of this study was to determine the optimal ASiR-V strength for lung analysis [88]. Images were acquired at 9.5 mGy (full dose) and 3mGy (low dose) and reconstructed with ASiR-V at different levels (0 to 100% every 20%) and a lung kernel. On the phantom, the image quality was assessed with an updated NPWE model observer to qualify the detectability (d' index). This index was compared to the ratings the radiologist obtained on patient acquisitions.

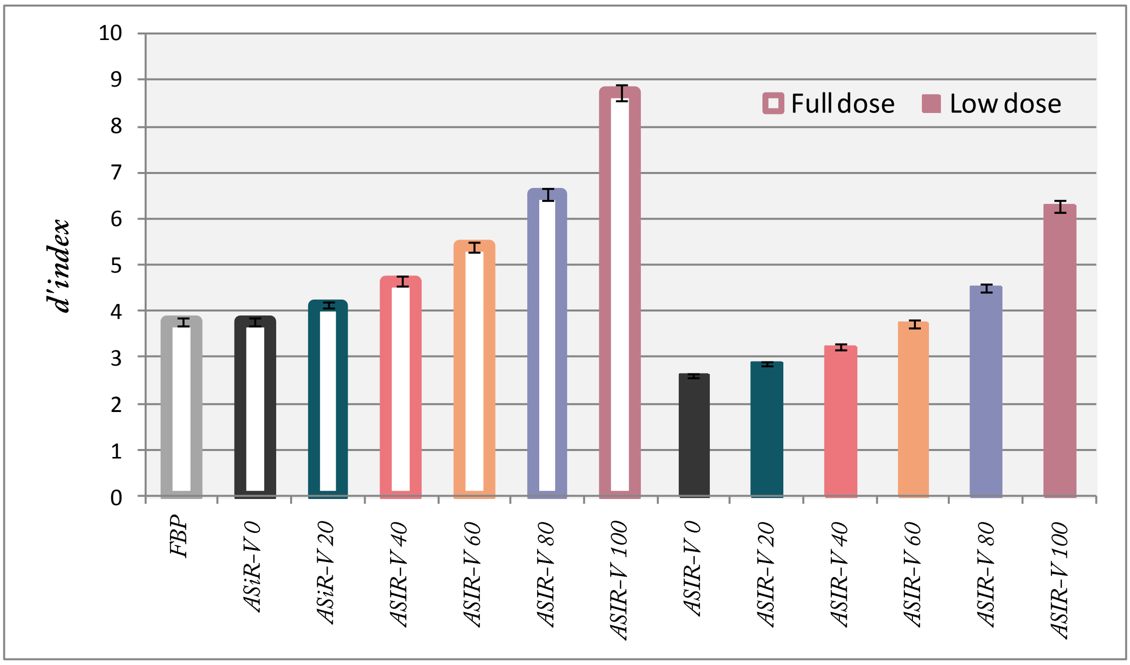

Figure 17 left: d' index calculated at full dose (7.5mGy) as a function of ASiR-V strength levels and at 9.5mGy for FBP algorithm. Right: d' index calculated at full dose (3.0mGy) as a function of ASiR-V strength levels and at 9.5mGy for FBP algorithm.

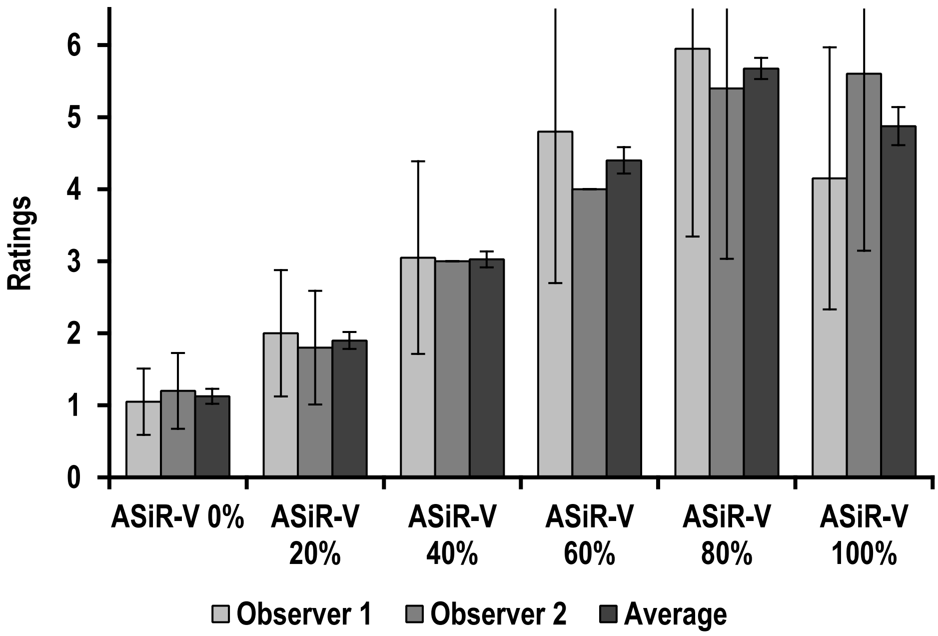

Figure 18: individual IQ scores of both raters and mean value

The results showed that the detectability index increased with the level of ASiR-V whatever the dose, and the maximum was obtained with ASiR-V 100% (Figure 17 left). Thus, with the low dose acquisition, the d' index with ASiR-V at 80 or 100% was higher than at full dose with FBP (Figure 17 right). For radiologists the best image quality score was obtained with ASiR-V at 80% or 100 % according to the readers (Figure 18).

With this kind of study the performance of model and human observers are not evaluated on the same images, but we found a correlation between the two kinds of observers. Prospectively, with phantom measurements, it was possible to find the right level of IR on clinical protocols before validating these results with human observer assessment on patient images.

8.2.2 Impact of the reconstruction plane on the image quality

This work was focused on comparing image quality in all three reconstruction planes (axial, coronal and sagittal) using objective assessment methods adapted to IR [89]. The acquired data sets were reconstructed in the axial, sagittal, and coronal planes, using a nominal slice thickness of 0.625 mm and four different reconstruction algorithms: the classical FBP, the ASiR at a percentage of 50 %, and the ASiR-V at a percentage of 50 %; with these three algorithms the GE bone kernel was used. Finally, the GE model-based iterative reconstruction "VEO" algorithm was also used. Images were reconstructed with VEO 2.0 that was only compatible with the standard kernel and VEO 3.0 with resolution preference (RP) 05 and RP 20. We also used an updated NPWE model observer to assess the image quality in the three reconstructed plane.
As expected, a full model based algorithm like VEO improved the detectability in comparison to the other algorithms (Figure 19).

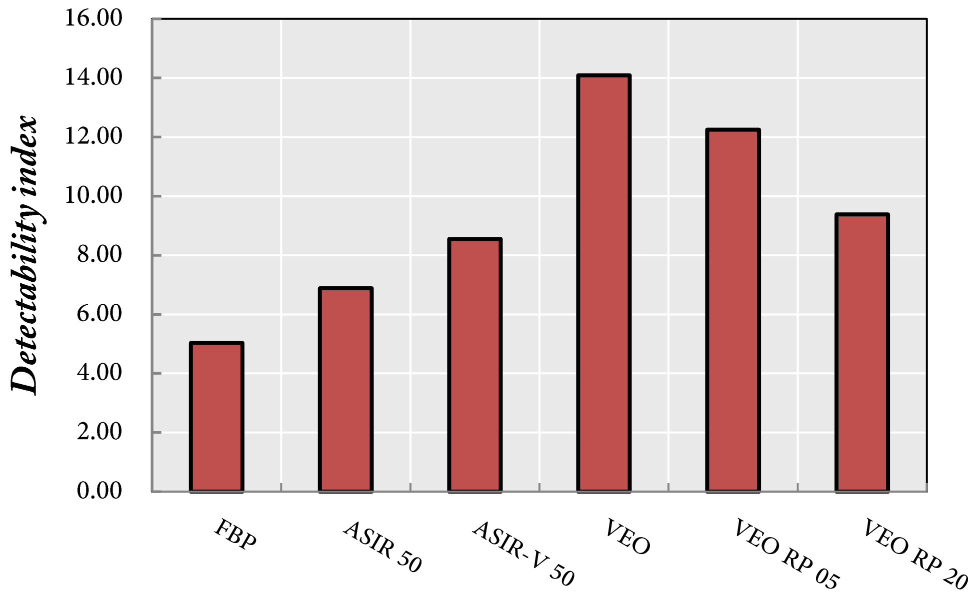

Figure 19: Detectability index obtained with the NPWE model in the axial plane with the differents algorithms

Moreover, major changes in the detectability were shown by the NPWE model observer in the sagittal and the coronal planes in comparison to the axial plane when images were reconstructed with FBP or statistical iterative algorithms (Figure 20).

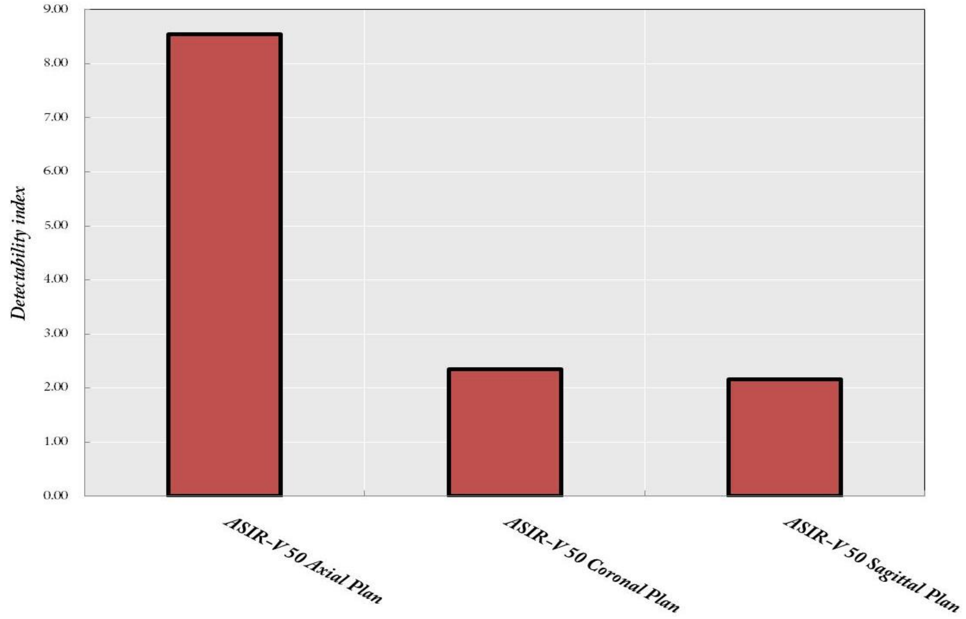

Figure 20: Detectability index for the ASiR-V algorithm for the different reconstructions planes for a lesion size of 1 mm

However, we observed a constant detectability in all reconstruction planes when using VEO, demonstrating that the use of this MBIR algorithm could help to improve diagnostic accuracy (Figure 21).

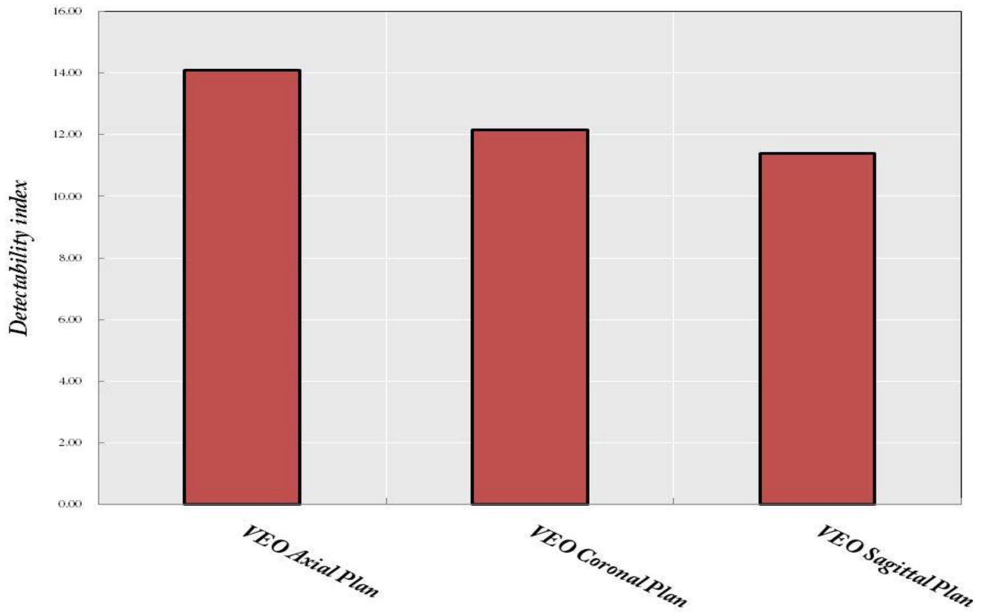

Figure 21: Detectability index for the VEO algorithm for the different reconstructions planes for a lesion size of 1 mm

This study indirectly impacts the clinical routine; indeed the majority of radiologists use the multi-planar reconstruction mode to make their diagnosis, and it is important to highlight that the image quality is not identical in the three planes.

8.3 Image quality in CT: A review

This work is a review that presents the different methods used to evaluate image quality in CT. First, the review explains the standard objective measurements of physical parameters, followed by a description of the methods usually used with human observer, and finishes with the clinically task-based approaches (i.e. model observer approach) that make the link between physical metrics and the human observer approach.

9 Scientific articles

Physical approach

Objective assessment of low-contrast detectability in computed tomography with Channelized Hotelling Observer

Racine Damien, Ba Alexandre H., Ott Julien G., Bochud François O., Verdun Francis R.

Phys Med. 2016 Jan;32(1):76-83. doi: 10.1016/j.ejmp.2015.09.011.

Objective task-based assessment of low-contrast detectability in iterative reconstruction

Racine Damien, Ba Alexandre H., Ott Julien G., Bochud François O., Verdun Francis R.

Radiat Prot Dosimetry February 27, 2016 doi:10.1093/rpd/ncw020.

Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners

Racine Damien, Viry Anaïs, Becce Fabio, Schmidt Sabine, Ba Alexandre, Bochud François O., Edyvean Sue, Schegerer Alexander, Verdun Francis R.

Med. Phys. **44**(9), e153–e163 (2017).

Benchmarking of CT for patient exposure opitmisation

Racine Damien, Ryckx Nick, Ba Alexandre, Ott Julien G., Bochud François O., Verdun Francis R.

Radiat Prot Dosimetry March 2, 2016 doi:10.1093/rpd/ncw021

Towards a standardization of image quality in abdominal CT: Results from a multicentre study

Racine Damien, Ryckx Nick, Ba Alexandre, Becce Fabio, Vïry Anais, Verdun Francis R. and Schmidt Sabine

Submitted in European Radiology

Clinical approach

Task-based assessment of impact of multiplanar reformations on objective image quality in iterative reconstruction in computed tomography

Racine Damien, Ott Julien G., Monnin Pascal, Omoumi Patrick, Verdun Francis R., Becce Fabio

Being processed for submission in Radiology

Review

Image quality in CT: From physical measurements to model observers

Verdun Francis R. and **Racine Damien**, Ott Julien G., Tapiovaara Markku J., Toroi Paula, Bochud François O., Veldkamp Wouter J., Schegerer Alexander, Bouwman Ramona W., Giron Irene H., Marshall Nicholas W., Edyvean Sue.

Phys Med. 2015 Dec;31(8):823-43. doi: 10.1016/j.ejmp.2015.08.007.

Physica Medica 32 (2016) 76–83

Contents lists available at ScienceDirect

Physica Medica

journal homepage: http://www.physicamedica.com

Original Paper

Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer

Damien Racine [a,b,*], Alexandre H. Ba [a], Julien G. Ott [a], François O. Bochud [a], Francis R. Verdun [a]

[a] Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland
[b] University Joseph Fourier Grenoble I, 621 Avenue Centrale, 38041 Saint-Martin-d'Hères, France

ARTICLE INFO

Article history:
Received 27 August 2014
Received in revised form 24 August 2015
Accepted 23 September 2015
Available online 26 October 2015

Keywords:
Computed tomography (CT)
Channelized Hotelling Observer model
Iterative reconstruction
Low contrast detectability

ABSTRACT

Purpose: Iterative algorithms introduce new challenges in the field of image quality assessment. The purpose of this study is to use a mathematical model to evaluate objectively the low contrast detectability in CT.

Materials and methods: A QRM 401 phantom containing 5 and 8 mm diameter spheres with a contrast level of 10 and 20 HU was used. The images were acquired at 120 kV with $CTDI_{vol}$ equal to 5, 10, 15, 20 mGy and reconstructed using the filtered back-projection (FBP), adaptive statistical iterative reconstruction 50% (ASIR 50%) and model-based iterative reconstruction (MBIR) algorithms. The model observer used is the Channelized Hotelling Observer (CHO). The channels are dense difference of Gaussian channels (D-DOG). The CHO performances were compared to the outcomes of six human observers having performed four alternative forced choice (4-AFC) tests.

Results: For the same $CTDI_{vol}$ level and according to CHO model, the MBIR algorithm gives the higher detectability index. The outcomes of human observers and results of CHO are highly correlated whatever the dose levels, the signals considered and the algorithms used when some noise is added to the CHO model. The Pearson coefficient between the human observers and the CHO is 0.93 for FBP and 0.98 for MBIR.

Conclusions: The human observers' performances can be predicted by the CHO model. This opens the way for proposing, in parallel to the standard dose report, the level of low contrast detectability expected. The introduction of iterative reconstruction requires such an approach to ensure that dose reduction does not impair diagnostics.

Introduction

Computed Tomography (CT) represents about 10% of all radiological examinations in Switzerland but about 70% of the collective effective dose [1]. Since no dose limit is applicable for patients, a first attempt to optimize patient exposure in radiology was the introduction of diagnostic reference levels (DRL) [2]. This concept makes it possible to identify situations in which dose level is beyond the examinations' common practices [3]. Despite this, a focus restricted to dose alone is highly insufficient in a framework of optimization between the radiological risk and image information. A surrogate to assess image information is the measurement of physical metrics such as image noise, spatial resolution, and contrast-to-noise ratio. However, these concepts are only valid for linear systems or algorithms. The introduction of iterative reconstructions into CT introduces new challenges in the field of image quality assessment since most of the standard metrics are no longer applicable [4–6]. State-of-the art medical image quality assessment takes another approach by defining image quality as how well the desired information for a given task can be extracted from an image [7]. Simple binary tasks such as discrimination between the presence and absence of pathology among a given population, are usually characterized by the use of Receiver Operating Curve (ROC) studies [8,9]. Unfortunately these studies are time-consuming [10] and difficult to implement in practice. Therefore, it is necessary to develop tools such as model observers [11,12] which make it possible to quantify image quality using a similar paradigm but in much simpler ways [13,14].

It has been shown that mathematical model observers, such as the Non Pre-whitening model with Eye-filter (NPWE) or Channelized Hotelling Observer (CHO) [15], can predict the capacity of human vision to detect low contrast targets. The advantage of this approach is that it enables testing the whole imaging chain [13] but it requires a substantial amount of data to be statistically robust

* Corresponding author. Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland. Tel.: +41 21 31 49 249; fax: +41 21 314 8299.
E-mail address: Damien.Racine@chuv.ch (D. Racine).

http://dx.doi.org/10.1016/j.ejmp.2015.09.011
1120-1797/

The purpose of this work was to investigate if the approach of image quality assessment by means of the CHO model observer could be applicable in routine practice in order to facilitate a control of image quality while reducing patient exposure. Thus, our work used a limited number of acquisitions with a specifically designed phantom that allowed the sampling of several realizations per slice. The use of such a methodology could then produce an image quality indicator that could be analyzed with the standard dose report. We also compared the results of the CHO model used with the outcome of human observers while varying several acquisition and image reconstruction parameters on an abdominal phantom.

Materials and methods

Data acquisitions

The QRM 401 phantom (Moehrendorf, Germany; see Fig. 1) was imaged on a GE 750 HD scanner (GE Healthcare, USA). The phantom is made of muscle, liver, spleen and bone (vertebrae) tissue equivalents at 120 kV. Two dedicated moduli (moduli A and B) can be inserted in the phantom shell. Modulus A includes spheres of various diameters: 8, 6, 5, 4 and 3 mm; each size having a contrast of 10 and 20 HU relative to the background at 120 kV. The phantom enables the assessment of in-plane and axial low-contrast detectability. In this study we only considered the in plane low contrast detectability for two sphere diameters: 5 and 8 mm. The spheres of each contrast level were positioned within the same slice position. Modulus B is homogeneous and was used to produce images free from low contrast target.

Acquisitions were performed at 120 kVp in helical mode (0.984 as pitch factor). In order to get $CTDI_{vol}$ values of 5, 10, 15 and 20 mGy the X-ray tube rotation time was kept constant [illegible] while varying the tube current. Indicated $CTDI_{vol}$ values were verified as described in the IEC 60601-2-44. The X-ray beam collimation geometry was 64×0.625 mm (leading to a total X-ray collimation at isocenter of 40 mm) and the reconstructed scan FOV was 320×320 mm using a 512×512 matrix size. Slices were reconstructed with a nominal thickness of 2.5 mm and a slice interval of 2.5 mm. The reconstruction filter used was the standard “body” filter provided by the manufacturer.

Images were reconstructed in the axial plane with three algorithms: filtered back-projection (FBP), and two iterative algorithms: Adaptive Statistical Iterative Reconstruction 50% (ASIR 50%) and the model-based iterative reconstruction (VEO 2.0) [16–20]. ASIR 50% was chosen since it corresponds to the option that is used for standard abdominal acquisitions in our center.

We investigated 48 different categories (3 reconstruction algorithms × 4 dose levels × 2 signal sizes × 2 contrast levels). The phantom with modulus A was (see Fig. 1) positioned at the isocenter of the CT unit and scanned ten times for each dose level, without changing its position between acquisitions. The phantom with modulus B was scanned only once for each dose level without changing its position between acquisitions.

Generating signal-absent and signal-present images

The program was implemented with the Python programming language. The first step performed by the software was the automatic production of ROIs. For that, the vertebra which represents the reference point was searched in the central slice. Using the coordinates of this reference point, a relative reference frame was created in the image and the ROIs were created automatically based on pre-established coordinates (derived from the technical plan of the phantom). The vertebra was used as a reference because it is the most contrasted material present in the phantom, ensuring that the template matching method is robust enough even at low dose levels. For each acquisition, 4 signal-centered ROIs per signal size/contrast combination (22×22 pixels; pixel size of 0.625 mm) were automatically extracted from the images. Signal-absent ROIs were extracted on images of Modulus B using the same in-plane (x,y) coordinates as the images obtained with Modulus A. However, the signal-absent ROIs were extracted in successive slices whereas the signal-present ROIs were extracted at a unique longitudinal position. In the following section ROIs will be called signal-present images if the signal is present in the ROIs or signal-absent images if the signal is absent. The sample consists of 40 signal-present cases (see Fig. 2) and 40 signal-absent cases (see Fig. 3) for each category. The same images were used for model and human observers.

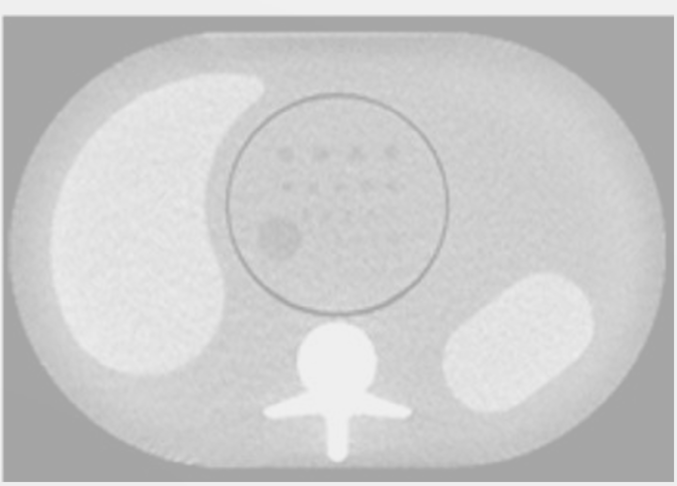

Figure 1. QRM 401 phantom acquired with MBIR at 20 mGy.

Figure 2. Signal-present image (sphere of 8 mm/20 HU).

Figure 3. Signal-absent image.

Human observer study

Six medical physicist students rated the images. These naive human observers (non radiologist) conducted four alternative forced choice (4-AFC) tests (see Fig. 4). The images were displayed on a Siemens SMM 21200P screen in conformity with DICOM 3.14 and AAPM TG18 standards [21]. The reading of images was performed in a room with an ambient light level of about 10 lux. Observers were at a distance of about 50 cm from the screen. No time limit was imposed on the observer to make their decision. During a 4-AFC study with 4 independent image combinations, three images with signal-absent and one with signal-present images were displayed and the observer was requested to identify which image contained the signal. To avoid any bias, the signal is randomly assigned to one of the 4 positions. The images were magnified to 128 × 128 pixels using a bilinear interpolation to appear at their actual size on the screen. The display window level and window width were adjusted to produce the best signal and then kept constant.

Each observer tested 30 images per category for a total of 1440 images. The test was divided into three sessions (480 trials per session) in order to minimize the loss of concentration; each session interval at least 24 hours. The first session began with a training session with 52 images acquired at high doses ($CTDI_{vol}$ = 35 and 50 mGy). During this session the result of each trial, "correct" or "incorrect", was shown to the observers after they replied.

For each category, at the end of the test each observer obtained a percentage of correct responses (PC). This metric represents the probability of correctly identifying the image containing the signal, and the higher the PC, the better the performance. PC was then converted into detectability to enable comparison between the performances of model observers with those of the human observers using Eq. (1) [15].

$$PC = \int d\lambda \varphi(\lambda - d_{4AFC}) \Phi(\lambda)^3 \tag{1}$$

where Φ is the cumulative Gaussian distribution function and d_{4AFC} represents the detectability obtained from a 4-AFC test performed by the human observers. The value of d_{4AFC} can be found using specific tabulated values [22–24].

Model observer: Channelized Hotelling Observer

Model observers are mathematical models based on the statistical decision theory to estimate the detection performance of ideal or human observers. In this study a linear anthropomorphic CHO model observer was chosen. The decision variable which is the outcome of the model is given by the dot product between the template $\mathbf{w}$ and the reconstructed image g_i (i = 0 or i = 1 respectively representing signal-absent or signal-present hypothesis), expressed as an N × 1 column vector (see Fig. 5) [7,13].

$$\lambda_i = \mathbf{w}^{T} \mathbf{g}_i \tag{2}$$

The CHO model is a derivative of the Hotelling Observer (HO) which is too computing expensive to be used in practice [9]. To reduce the dimensionality of HO and take advantage of the spatial selectivity behavior of the human visual system, the image is first passed through J channels where J is significantly lower than N [25]. The channel output, a scalar v_i, is obtained by the dot product between the channel u_j and the image g.

$$\mathbf{v}_i = (\mathbf{u}_j)^{T} \mathbf{g} \tag{3}$$

Thus, U, the matrix representation of the channel filters, is an N × J matrix where each column is one of channels [26].

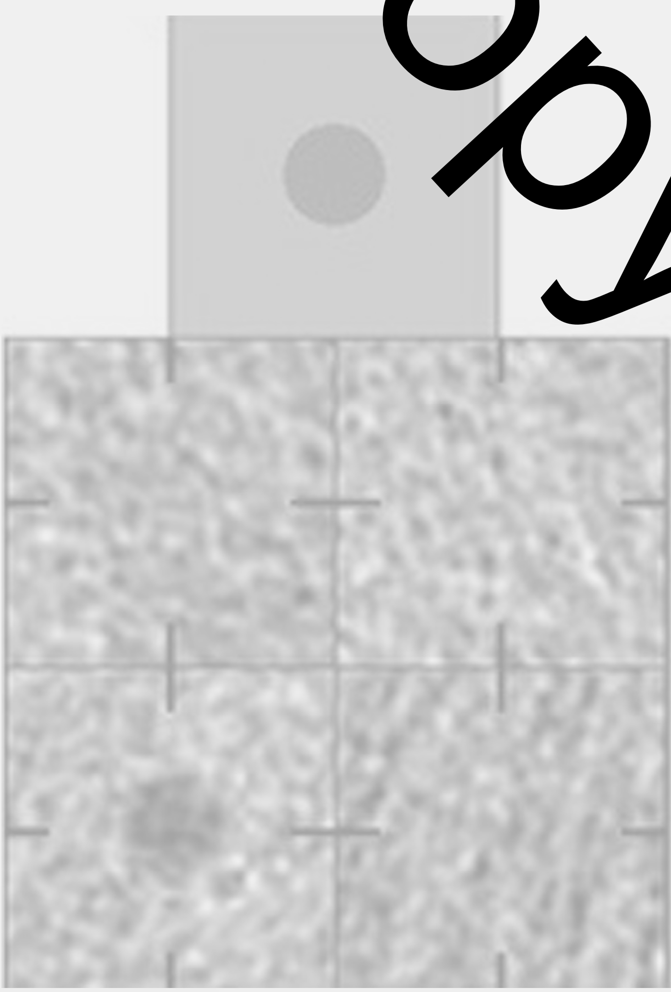

Figure 4. Example of test 4 AFC.

$$g(x,y) = \begin{bmatrix} g_{0,0} & g_{0,1} & \cdots & g_{0,m} \\ g_{1,0} & g_{1,1} & \cdots & g_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n,0} & g_{n,1} & \cdots & g_{n,m} \end{bmatrix} \rightarrow \mathbf{g} = \begin{bmatrix} g_{0,0} \\ g_{0,1} \\ g_{0,m} \\ g_{1,0} \\ g_{1,1} \\ \vdots \\ \vdots \\ g_{n,m} \end{bmatrix}$$

$$w(x,y) = \begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,m} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,0} & w_{n,1} & \cdots & w_{n,m} \end{bmatrix} \rightarrow \mathbf{w} = \begin{bmatrix} w_{0,0} \\ w_{0,1} \\ w_{0,m} \\ w_{1,0} \\ w_{1,1} \\ \vdots \\ \vdots \\ w_{n,m} \end{bmatrix}$$

Figure 5. Image and template as column vectors.

$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_j]$

For the type of targets to be detected in this study the CHO model, using the dense of difference of Gaussian (D-DOG) channels type, is recognized as a good model of human vision, and this is even with a limited number of 10 channels enabling a drastic reduction of the images required to compute the model observer outcome [27,28].

In this model, the radial profile of each frequency is given by the following formula:

$$C_j(\rho) = e^{-\frac{1}{2}\left(\frac{\rho}{Q\sigma_j}\right)^2} - e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2} \qquad (4)$$

where ρ is the spatial frequency and σ_j is the standard deviation of each channel. Each σ_j value is given by $\sigma_j = $ [illegible]. Factor Q is the bandwidth of the filter. Generally the parameters used are [illegible] = 0.0[illegible], [illegible] = 1.4 and Q = 1.[illegible] values from [27].

The template of the resulting covariance matrix was calculated from all images containing no signal according to:

$$\mathbf{w}_{CHO} = (\mathbf{K}_{v/n}) [illegible] \mathbf{v}_n\rangle \qquad (5)$$

where $(\mathbf{K}_{v/n}) = \boldsymbol{U}^T \mathbf{K}_{v/n} \boldsymbol{U}$.

$\mathbf{K}_{v/n}$ represents the covariance matrix computed from channelized images containing no signal. In Eq. (5) $\mathbf{v}_s$ represents the vector containing the data of the signal images as seen through the channels, and $\mathbf{v}_n$ represents the vector containing the data of the signal-absent as seen through the channels [illegible].

The decision variable of the CHO model is obtained by combining Eq. (5) and Eq. (2). However, with the CHO model the decision variable is not computed with the images but with the channelized images (v_i):

$$\lambda_{CHO} = \mathbf{w}^T{}_{CHO} \mathbf{v}_i \qquad (6)$$

Internal noise

Model observers like CHO with well suited channels are more efficient than human observers for simple detection tasks [illegible] Signal Known Exactly/Background Known Exactly (SKE/BKE). To adjust the detection outcomes of model observers to human observers it is necessary to add some internal noise, ε, to the model observer formula as shown by Eq. (7) [30].

$$\lambda_{noisy} = \lambda + \varepsilon \qquad (7)$$

Internal noise ε is added to the decision variable λ, with a probability proportional to the standard deviation of the distribution of the decision variable from the signal absent images [30].

$$\varepsilon = \alpha \times \sigma_{bg} \times \xi \qquad (8)$$

where α is the weighting factor and ξ is a random number generated between −1 and 1, σ_{bg} is the standard deviation of the distribution of the decision variable of signal-absent.

In this study, the coefficient **α** was calibrated using images containing the signal 8 mm/10 HU at 15 mGy reconstructed with FBP and VEO. **α** was varied between 0 and 10 iteratively. The value that minimized the difference between the model observers and the human observer outcomes for each algorithm was then selected.

Assessment of the model outcomes

For each category, 560 decision variables were calculated (520 from signal-absent images and 40 from signal-present images). The ROC curves were then generated from pairs of TPF and FPF and the area under the curve was calculated by the trapezoidal method using 100 points.

Concerning the uncertainties of the results the average and standard deviation of the area under the curve (AUC) are obtained using the bootstrap method [31]; in our study the error bars represent plus or minus one standard deviation (68% for a Gaussian distribution). In order to estimate the mean and the standard deviation, the bootstrap was made using 1000 iterations for each category, and for each iteration, 520 signal-absent images and 40 signal-present images were randomly selected and replaced. Finally, to compare the performance of the CHO and the human performance, the AUC and its uncertainties were converted into detectability index (d_A) using Eq. (9) to be used as a figure of merit [15]. For our calculations, the maximum value was set to 6 whereas theoretically, detectability varies between 0 and infinity. Obviously, the higher is the index value, the better is the signal visibility.

$$d_A = 2\Phi^{-1}(2AUC - 1) \qquad (9)$$

where Φ is the normal cumulative distribution function.

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{(-y)^2} dy \qquad (10)$$

Results

Comparison of FBP, ASIR 50% and MBIR

In this section, algorithms FBP, ASIR 50% and MBIR are compared with the CHO model without internal noise implemented with a D-DOG. Figure [illegible] shows the performance of the CHO model without noise addition for the algorithms. FBP is taken as the reference algorithm. The detectability index d_A, varies from 0.67 to 6; i.e. it varies from almost no detection to a perfect detection. The results for ASIR 50% are comparable to the results of FBP whatever the size and contrast tested. ASIR 50% is however slightly better than FBP when the dose reaches a certain level. For targets 5 mm/10 HU, the performances between the algorithms are very similar. When the signal is hardly detectable MBIR does not improve the performance. Nevertheless, with 8 mm/10 HU or [illegible]m/20 HU, MBIR improves the detectability resulting in better outcomes whatever the dose when compared to ASIR 50% and FBP. The detectability index d_A, generally increases with dose reaching a plateau. This plateau is reached faster with MBIR algorithm than with ASIR 50% or FBP.

Internal noise calibration

Figure 7 shows the variation of d_A as a function of α at a dose level of 15 mGy for the sphere 8 mm/10 HU reconstructed with FBP. As expected, the higher the α, the lower the d_A. From these data it appears that α set to 3.[illegible] provides a good match between the CHO and human observers for the image of this category. We decided to take that value for the other categories. The calibration was also performed with the algorithm VEO using the category 8 mm/10 HU at 10 mGy (see Fig. 8). The resulting alpha coefficient is 3.6 which is quite similar to the previous one but it enables a better comparison.

Correlation of performances between model observers and human observers

All the detectability indexes obtained using the CHO model were compared with the 4-AFC results for images reconstructed using the algorithms MBIR and FBP, only since ASIR 50% led to comparable results

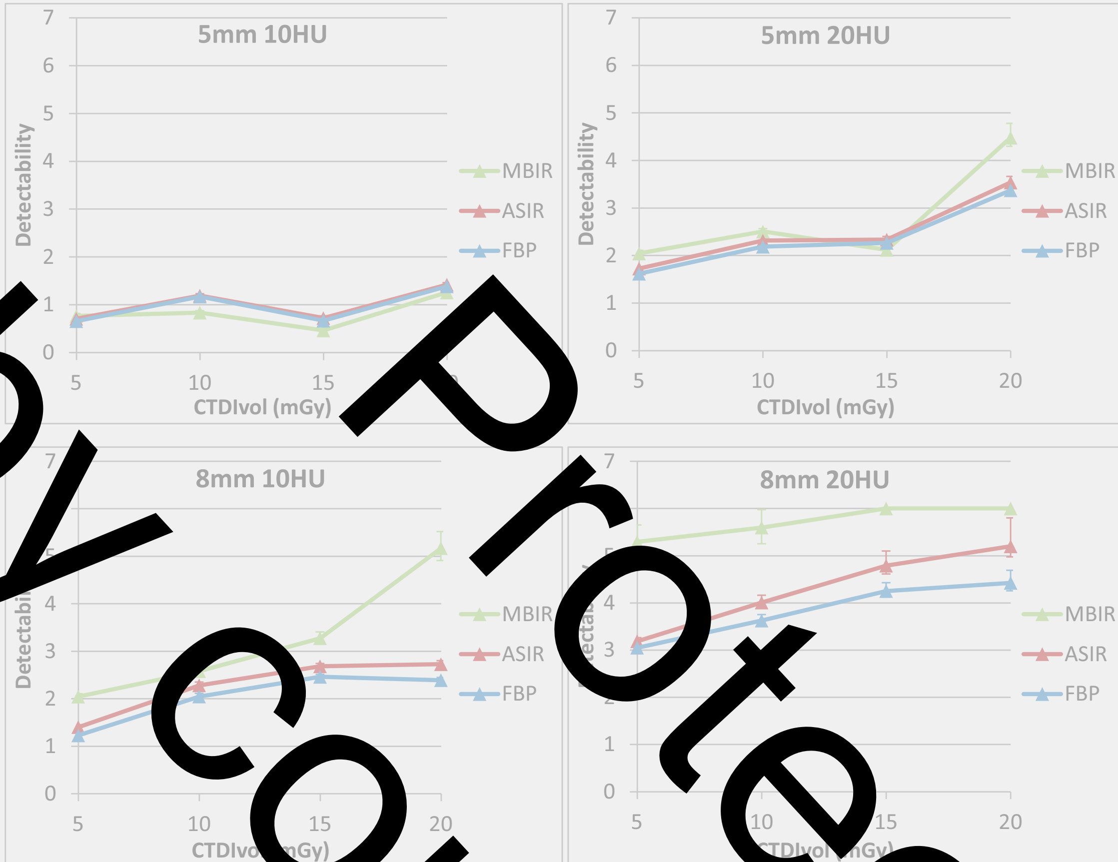

Figure 6. Comparison between FBP, ASIR 50% and MBIR algorithms.

than FBP. For each dose level, performances were strongly correlated except for the 8 mm/20 HU at 5 mGy. The Pearson coefficient is [illegible] for the 5 mm sphere FBP-reconstructed, 0.948 for 8 mm sphere FBP-reconstructed, 0.971 for the 5 mm sphere MBIR-reconstructed and [illegible] for 8 mm sphere MBIR-reconstructed. Error bars for the model observer represent plus or minus one standard deviation uncertainty obtained from 200 internal noise realizations. For human observers, uncertainty is also plus or minus one standard deviation for responses recorded during the 4-AFC experiments.

All human observer results show a dose dependent increase in the detectability index. This is also the case when the size and contrast increase. For 8 mm/20 HU the increase in dose is not associated

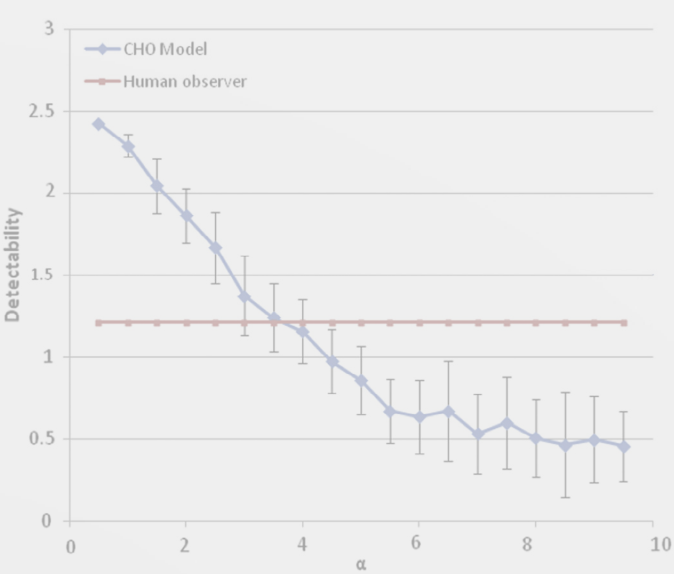

Figure 7. Internal noise calibration for an 8 mm and 10 HU sphere reconstructed with FBP algorithm.

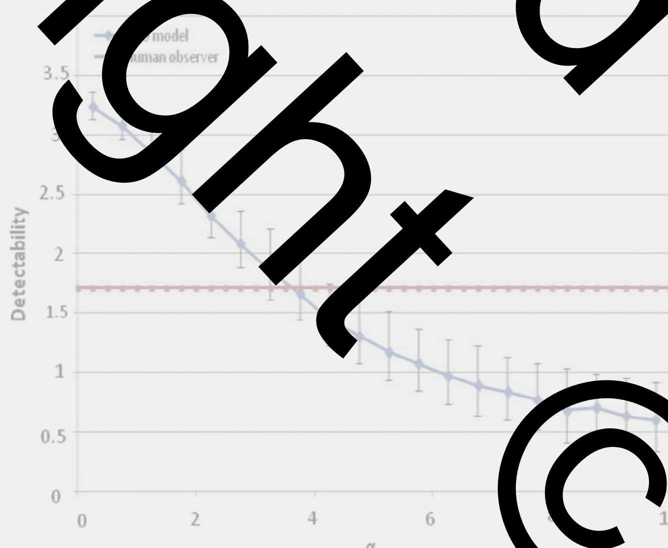

Figure 8. Internal noise calibration for a 8 mm and 10 HU sphere reconstructed with MBIR algorithm.

Figure 9. Performance comparison between the CHO model and human observers.

with a detectability index benefit since human observer outcomes are already very good (see Fig. 9).

To compare the performances between CHO (internal noise added) and humans, the Bland–Altman plot was used. Its ordinate is the difference between the values obtained by the two types of observers and its abscissa is the average value of the detectability index obtained by the two types of observers. Each point represents a class and the dotted lines represent plus or minus two standard deviations (95% for a Gaussian distribution). All the points are in the range plus or minus two standard deviations (see Fig. 10). Internal noise provides a good agreement between the responses of human observers and model observers. Also there is no relationship between the deviations from the mean and increased detectability.

Discussion

Currently, patient radiation protection is a major issue and there is a tendency to significantly reduce dose without paying much attention on the potential loss of low contrast detectability. If this parameter was controlled by CNR measurements in the past (using FBP and keeping the same reconstruction kernel) the introduction of iterative reconstruction algorithms does not allow such an approach anymore. One way to ensure the detection of a low contrasted lesion is to try to mimic human detection with mathematical model observers such as the CHO model used in this study. Our results show that the trends of low contrast detection provided with the CHO model used are compatible with human observers. However, without the addition of internal noise the CHO model outperforms human outcomes. It is therefore necessary to use the internal noise of the CHO model to obtain a good correlation between the responses. We have shown that under our conditions a unique additional noise setting gave satisfactory results whatever the sizes, contrasts and dose involved. Thus, our method makes it possible to link a dose level to low contrast detectability performances. This information should improve the way optimization between image quality and patient exposure is balanced.

According to manufacturers, iterative reconstructions enable drastic dose reduction without major loss of image quality. Our results show that in terms of low contrast detectability caution must be exercised in particular with the iterative reconstruction of the first generation tested (ASIR 50%) in spite of having [illegible] a percentage recommended to get an image quality improvement without

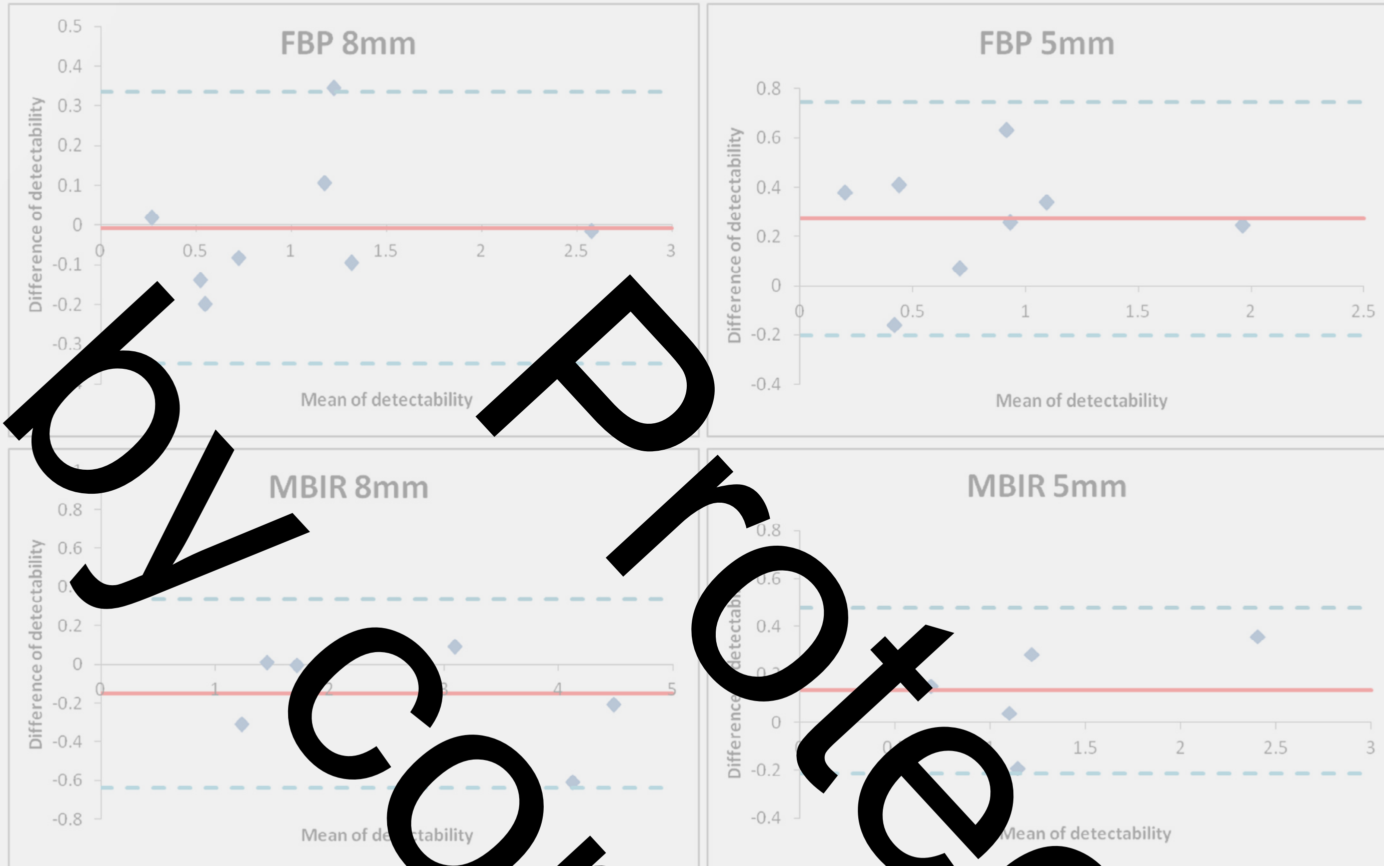

Figure 10. Bland–Altman plot of detectability index difference between human and model observers.

major image texture changes [32]. Concerning the model based iterative reconstruction (MBIR), in spite of being very computing extensive, low contrast detectability cannot be recovered at a low dose and for a very low contrast level. However, after a certain increase in dose the use of MBIR leads to much better results in terms of low contrast detectability than FBP or ASIR 50%. This kind of information is important when willing to lower patient exposure.

One limitation of our study when willing to calibrate the mathematical model observer with human results is the design of the phantom. It enables getting four spheres of a given size and contrast level per acquisition which is an advantage, but these spheres are very close to each other which require the use of a pixel interpolation to get a reasonable image size to be presented to the human observer. For such calibration purposes one should avoid placing several spheres within the same slice in order to generate large ROIs compatible with the suggestions of Yu [29] who proposed an ROI size of 4–5 times the size of the signal. Moreover the CT iterative reconstruction is not shift invariant, so to have a maximum of spheres in the minimum of space the phantom used in this project provides 4 spherical ROIs per acquisition. Unfortunately for this compromise, which is nonetheless an advantage in terms of being able to use this protocol to evaluate clinical protocols or CT units, it was necessary to create the signals close to each other which might also introduce some correlations from one signal affecting the values of nearby signals.

Finally, our results apply in a simple situation in comparison to the actual environment. The background images are homogeneous and the task is quite basic. However, we have been able to demonstrate that dose reduction must be introduced while keeping in mind that the detection of low contrast structures might be lost. In such a situation some kind of information should be displayed on the unit to inform the radiologists about the type of low contrast sphere they will not be able to detect.

Conclusion

CHO model coupled to D-DOG channels can be used to predict human observer performance for a 4AFC even with a limited number of acquisitions compatible with routine quality control measurements in order to assess the low contrast detectability of the acquisition for the FBP and iterative algorithms. From our results, we can conclude that the model based iterative generation algorithm (MBIR) offer superior image quality than FBP or ASIR 50% at equivalent dose. Thus, MBIR certainly offers a potential for dose reduction.

A CHO model, such as the one used in our study, could be used in routine to qualify the image quality for any given acquisition protocol. The results provided are easy to present and can be well understood by radiologists and radiographers. Finally, the use of such model observers appears to be necessary to avoid dose reduction that would significantly impair low contrast detectability.

Acknowledgments

I would like to thank Christel Elandoy for his active participation in acquisitions, but also Ivan Diaz, Archonteia Kyroudi and Nick Ryckx for their participation in detection studies and Georg Kropat for his valuable statistical advice.

This work was partly supported by a grant from the Swiss National Science Foundation (SNSF) [No. 320030-140995].

References

[1] Samara ET, Aroua A, Bochud FO, Ott B, Theiler T, Treier R, et al. Exposure of the Swiss population by medical x-rays: 2008 review. Health Phys 2012;102:263–70. doi:10.1097/HP.0b013e31823513ff.

[2] Office fédéral de la santé publique OFSP. Notice R-06-06: niveaux de référence diagnostiques en tomodensitométrie 2010.

[3] Smith-Bindman R. Is computed tomography safe. N Engl J Med 2010;363:1–4.

[4] Silva AC, Lawder HJ, Hara A, Kujak J, Pavlicek W. Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. AJR Am J Roentgenol 2010;194:191–9. doi:10.2214/AJR.09.2953.

[5] Miéville FA, Gudinchet F, Brunelle F, Bochud FO, Verdun FR. Iterative reconstruction methods in two different MDCT scanners: physical m[illegible]trics and 4-alter[illegible]e forced-choice detectability experiments – a phant[illegible]roach. Phy[illegible]d 2013;29:99–110. doi:10.1016/j.ejmp.2011.12.004.

[6] V[illegible]nav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Obje[illegible] assessm[illegible] of [illegible]ality and dose reduction in CT iterative reco[illegible]tion. Med P[illegible] 2014;41[illegible]1904. doi:10.1118/1.4881148.

[7] Barrett H[illegible] Myers KJ. Foundations of image scie[illegible] Hoboke[illegible] New Jers[illegible] Wiley-In[illegible]science; 2004.

[8] Intern[illegible]nal C[illegible]mission on Radiation Units and Measurements. Receiver [illegible]g cha[illegible]eristic analysis in medical imaging. ICRU Rep. No. 79, vol. 79 [illegible]ethesda ([illegible]): International Commission on Radiation Units [illegible] Measureme[illegible]; 2008.

[9] Barrett HH[illegible]o J, Rol[illegible], Myers KJ. Model observers for assessmen[illegible] image quality. [illegible] N[illegible]ad Sci U S A 1993;90:9758–65.

[10] [illegible]ang [illegible], Stayman JW, Tward DJ, Zbijewski W, Prince JL, et al. Analysis of [illegible] domain task-based detectability index in tomosynthesis and cone-bea[illegible] CT in relation to human observer performance. Med Phys 2011;38:1754–6[illegible] doi:10.1118/1.3560428.

[11] He X, Park S. Model obser[illegible]n med[illegible] imaging research. Theranostic[illegible] 2013;3:774–86. doi:10.71[illegible]nno.5138.

[12] Burgess AE, Jacobson FL, Ju[illegible] PF. Human o[illegible]erver detection experiments with mammograms and power[illegible]w noise. Med Phys 2001;28:419–37.

[13] Hernandez-Giron I, Geleijns [illegible]alzado A, Veld[illegible]H. Automated assessment of low contrast sensitivity fo[illegible]tem[illegible]ing a m[illegible]el observer. Med Phys 2011;38:S25–35. doi:10.1118/1.35777[illegible]

[14] COCIR. CT Manufacturer's Voluntary Co[illegible]mitment Rega[illegible]ng CT Dose to HERCA Working Group 2013.

[15] Beutel J, Kundel H, Van Metter R. Handb[illegible]k of me[illegible] imagi[illegible] vol. I. Physics and psychophysics. Bellingham, Washingt[illegible]ress; [illegible]95–695.

[16] Singh S, Kalra MK, Hsieh J, Licato PE, Do S, Pien HH[illegible]al. Abdo[illegible]inal CT: comparison of adaptive statistical iterative and [illegible]ered back p[illegible]ection reconstruction techniques. Radiology 2010;[illegible]73–83. doi[illegible].1148/radiol.10092212.

[17] Hara AK, Paden RG, Silva AC, Kujak JL, La[illegible]der HJ[illegible]. Itera[illegible]e reconstruction technique for reducing body radiation d[illegible] CT: feasi[illegible]ty study. AJR Am J Roentgenol 2009;193:764–71. doi:10.2214/AJR.09.239[illegible]

[18] Patino M, Fuentes JM, Singh S, Hahn PF, Sahani DV. Iterative reconstruction techniques in abdominopelvic CT: technical concepts and clinical implementation. AJR Am J Roentgenol 2015;205:W19–31. doi:10.2214/AJR.14.13402.

[19] Thibault J-B, Sauer KD, Bouman CA, Hsieh J. A three-dimensional statistical approach to improved image quality for multislice helical CT. Med Phys 2007;34:4526–44. doi:10.1118/1.2789499.

[20] Scheffel H, Stolzmann P, Schlett CL, Engel L-C, Major GP, Károlyi M, et al. Coronary artery plaques: cardiac CT with model-based and adaptive-statistical iterative reconstruction technique. Eur J Radiol 2012;81:e363–9. doi:10.1016/j.ejrad.2011.11.051.

[21] Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. Med Phys 2005;32:1205–25. doi:10.1118/1.1861159.

[22] Craven BJ. A table of d' for M-alternative odd-man-out forced-choice procedures. Percept Psychophys 1992;51:379–85.

[23] Hacker MJ, Ratcliff R. A revised table of d' for M-alternative forced choice. Percept Psychophys 1979;26:168–70. doi:10.3758/BF03208311.

[24] Dahlquist G, Björck Å. Numerical Methods. Courier Corporation; 2012.

[25] Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. J Opt Soc Am A 1987;4:2447–57. doi:10.1364/JOSAA.4.002447.

[26] Baydush AH, Catarious DM, Lo JY, Floyd CE. Incorporation of a Laguerre–Gauss channelized hotelling observer for false-positive reduction in a mammographic mass CAD system. J Digit Imaging 2007;20:196–202. doi:10.1007/s10278-007-9009-8.

[illegible]bbey CK, Barrett HH. Human- and model-observer performance in ramp-sp[illegible]trum noise: effects of regularization and object variability. J Opt Soc Am A 2[illegible]1;18:473–88. doi:10.1364/JOSAA.18.000473.

[28] Tse[illegible] H-W, Fan J, Kupinski MA, Sainath P, Hsieh J. Assessing image quality and d[illegible] redu[illegible] a new x-ray computed tomography iterative reconstruction [illegible]lgorithm[illegible]g model observers. Med Phys 2014;41:071910. doi:10.1118/1.4881[illegible]

[29] Yu [illegible]ng S, Chen L, Kofler JM, Carter RE, McCollough CH. Prediction of h[illegible]n observ[illegible]rformance in a 2-alternative forced choice low-contrast [illegible]ection [illegible]using [illegible]hannelized Hotelling observer: impact of radiation [illegible] and [illegible]tructio[illegible]lgorithms. Med Phys 2013;40:041908. doi:10.1118/1.47944[illegible]

[30] Leng S, [illegible]u L, Zha[illegible] Y, [illegible]arter R, Toledano AY, McCollough CH. Correlation between [illegible]odel obse[illegible] and human observer performance in CT imaging when lesion loca[illegible]i[illegible]ncertain[illegible] 2013;40:081908. doi:10.1118/1.4812430.

[31] Efron B, Tibshirani RJ. A[illegible]ntroduct[illegible] to the bootstrap. Boca Raton, Florida: CRC Press; 1994.

[32] Schindera ST, Odedra [illegible]Raza SA, Kim [illegible], Jang H[illegible]ucs-Farkas Z, et al. Iterative reconstruction algo[illegible]hm for CT: can ra[illegible]on dose be decreased while low-contrast detecta[illegible]lity is preserv[illegible]adiology 2013;269:511–18. doi:10.1148/radiol.13122[illegible]

Radiation Protection Dosimetry (2016), Vol. 169, No. 1–4, pp. 73–77
Advance Access publication 27 February 2016 doi:10.1093/rpd/ncw020

OBJECTIVE TASK-BASED ASSESSMENT OF LOW-CONTRAST DETECTABILITY IN ITERATIVE RECONSTRUCTION

Damien Racine*, Julien G. Ott, Alexandre Ba, Nick Ryckx, François O. Bochud and Francis R. Verdun
Institute of Radiation Physics, CHUV, Lausanne, Switzerland

*Corresponding author: damien.racine@chuv.ch

Evaluating image quality by using receiver operating characteristic studies is time consuming and difficult to implement. This work assesses a new iterative algorithm using a channelised Hotelling observer (CHO). For this purpose, an anthropomorphic abdomen phantom with spheres of various sizes and contrasts was scanned at 3 volume computed tomography dose index ($CTDI_{vol}$) levels on a GE Revolution CT. Images were reconstructed using the iterative reconstruction method adaptive statistical iterative reconstruction-V (ASIR-V) at ASIR-V 0, 50 and 70 % and assessed by applying a CHO with dense difference of Gaussian and internal noise. Both CHO and human observers (HO) were compared based on a four-alternative forced-choice experiment, using the percentage correct as a figure of merit. The results showed accordance between CHO and HO. Moreover, an improvement in the low-contrast detection was observed when switching from ASIR-V 0 to 50 %. The results underpin the finding that ASIR-V allows dose reduction.

INTRODUCTION

Image quality in computed tomography (CT), and hence clinical protocol optimisation, is a challenge because CT delivers the highest dose to the population in Switzerland, as in most Western countries[1]. To overcome this problem, manufacturers developed new technologies (e.g. iterative reconstruction) to minimise the amount of dose delivered. However, these new technologies must be evaluated and optimised[2,3]. To optimise protocols, it is possible to evaluate simple binary tasks, such as the discrimination between the presence and absence of a pathology, which can be characterised by the use of receiver operating characteristic (ROC) studies[4]. These methods can be used to assess image quality especially when new reconstruction algorithms are implemented[5]. Unfortunately, these studies are time consuming and difficult to implement. Developing a tool based on mathematical model observers like the channelised Hotelling observer (CHO) to optimise clinical protocols and ensure that dose reductions do not affect the detection of low-contrast structures (using a similar paradigm as ROC studies) is an effective way to perform CT image quality—dose optimisation[6]. Nowadays, a new algorithm is often assessed by model observer studies in the image domain because the new iterative algorithms do not respect the shift-invariance and cyclostationarity assumptions necessary to use Fourier-based metrics[7–9]. In this study, the CHO model that mimics the human observer performance was used to assess the dose reduction potential of a new iterative reconstruction algorithm while keeping the image quality constant.

MATERIALS AND METHODS

Data acquisition

An anthropomorphic abdomen phantom (QRM 401, Moehrendorf, Germany) simulating the attenuation produced by a thin patient (equivalent diameter 24 cm) was scanned on a new Revolution CT (GE Healthcare, USA). Two modules can be inserted in the middle: a homogeneous module similar to the phantom's background and a low-contrast module that contains 6 and 8 mm spherical targets with contrast levels of 10 and 20 HU at 120 kV.

The phantom was scanned using the helical mode with a pitch of 0.984 at 120 kVp. The tube current was adapted for three dose levels [5, 10 and 15 mGy, volume computed tomography dose index ($CTDI_{vol}$) values were calculated as described in the IEC 60601-2-44[10] in a 32-cm $CTDI_{vol}$ abdomen phantom]. However, the $CTDI_{vol}$ indicated was overestimated, because it was calculated for a 32-cm CTDI phantom, whereas the phantom measured only 24 cm.

Fifty acquisitions per condition were made, and images were reconstructed using a 320-mm display field of view (DFOV), a 512 × 512 matrix size and a reconstructed slice thickness of [illegible] mm associated to a reconstruction interval of [illegible] mm. A new iterative algorithm, adaptive statistical iterative reconstruction (ASIR-V), at different levels (0, 50 and 70 %), was used to reconstruct the images[11]. This algorithm is a mixture between ASIR and MBIR, but to decrease the time reconstruction, ASIR-V does not model system optics[12–15]. The ASIR-V level represents the noise reduction rate.

A total of 36 different categories were obtained (three dose levels, two ASIR-V levels, two contrast levels and two target sizes), and 200 regions of interest (ROIs) of 28 × 28 pixels containing centred signals were extracted for each category. The ROIs containing the noise were extracted from the homogeneous modulus at the same (x,y) location as the signals in order to avoid the problem of non-stationary noise.

Model observer: channelised Hotelling observer

In this work, three human observers (medical physicists) and the CHO model performed a four-alternative forced-choice (4-AFC) experiment in which a series of four images were displayed but only one contained a signal. The CHO is a linear anthropomorphic model that can be used to evaluate the image quality[2, 3]. The methodology used in this paper for the CHO is directly inspired by the methodology used by Ott *et al.*[16]. (The reader interested in all the mathematical details of the process will find extensive details in the chapter on the CHO model observer in Ott *et al.*).

For each trial of four images, the observer had to identify the image that was the most likely to contain the signal. The model observer compared the decision variable of the four displayed images and chose the one with the highest value as signal-present image. In the end, a percentage correct (PC) was calculated and used as a figure of merit.

The average and standard deviation of the model observer's PC were estimated by performing a bootstrap method[17]. In practice, 1000 4-AFC experiments were performed for each category and each 4-AFC experiment was created from 1[illegible] trials (1 image containing the signal and 3 signal-absent images per trial). Then, the mean and standard deviation of the 1000 PC values were computed. For human observers, the mean value and its standard deviation were calculated using the PC of the three observers. Error bars represent plus or minus one standard deviation.

Internal noise

To adjust the model's response to human observers' responses, it was necessary to reduce the model's performances with an internal noise ϵ because the CHO model with DDoG channels overestimates the performance of human observers in some conditions[18]. First, the decision variable λ was calculated as described in the methodology presented by Ott *et al.*[16], but at the end a random variable was added to obtain a noisy decision variable[8, 16]:

$$\lambda_{noisy} = \lambda + \varepsilon. \tag{1}$$

Internal noise ϵ was added to the decision variable λ with probability proportional to the standard deviation of the distribution of the decision variable amplitude images when the signal was absent:

$$\varepsilon = \alpha \times \sigma_{bg} \times \xi, \tag{2}$$

where α is the weighting factor, ξ is a random number generated between -1 and 1, and σ_{bg} is the standard deviation of the distribution of the decision variable of signal absent.

The α value was obtained by a calibration using images containing the signal 6 mm/20 HU at 10 mGy reconstructed with ASIR-V 50 %. The α value that minimised the difference between the model observers and the human observer was selected.

RESULTS

Figure 1 shows the humans' performances and Figure 2 the model observer's performances using ASIR-V algorithm at different levels. The category 8 mm/20 HU was too trivial (PC always equal to 1 in every condition), and it was used as a training test for each dose and algorithm level.

For human observers, the results suggest that the detectability slightly increases with the level of ASIR-V (Figure 1). Thus, when the dose increases, PC tends to reach a plateau. This phenomenon was similar when the size or the contrast levels increased. In some conditions, using a higher ASIR-V level slightly improved the detectability, especially at low contrast.

As can be seen in Figure 2, the trends obtained from the CHO with DDoG and without adding internal noise were very similar to those obtained with human observers, even if the model overestimates the human performance and clears the difference between ASIR-V levels. In the high dose range, a significant difference appeared when using ASIR-V 70 % instead of ASIR-V 50 %.

Figure 3 shows the variation of PC as a function of A at a dose level of 10 mGy for the signal 6 mm/20 HU reconstructed with ASIR-V 0 %. As expected, the higher the A, the lower the PC. The A value that provided the best match between the CHO performance and the human observers' performance for this category was chosen for other categories. The A value that provided the best match was equal to 4.0.

Adding internal noise within the CHO decreased its performance in order to match the human observers. This match was quantified with the root-mean-square error (RMSE). The value of the RMSE for all categories decreased from 0.62 to [illegible]78 when internal noise was added. Like the human performance, the PC increases with the dose when an internal noise is added. The human's performance is linked with the ASIR-V levels. However, for the model observer, the hierarchy between different ASIR-V levels is not very clear. For example, the PC for the category 8 mm/[illegible] HU became unexpectedly worse at 15 mGy. Thus, the image quality tends to a plateau between 10 and 1[illegible] mGy with ASIR-V 70 %

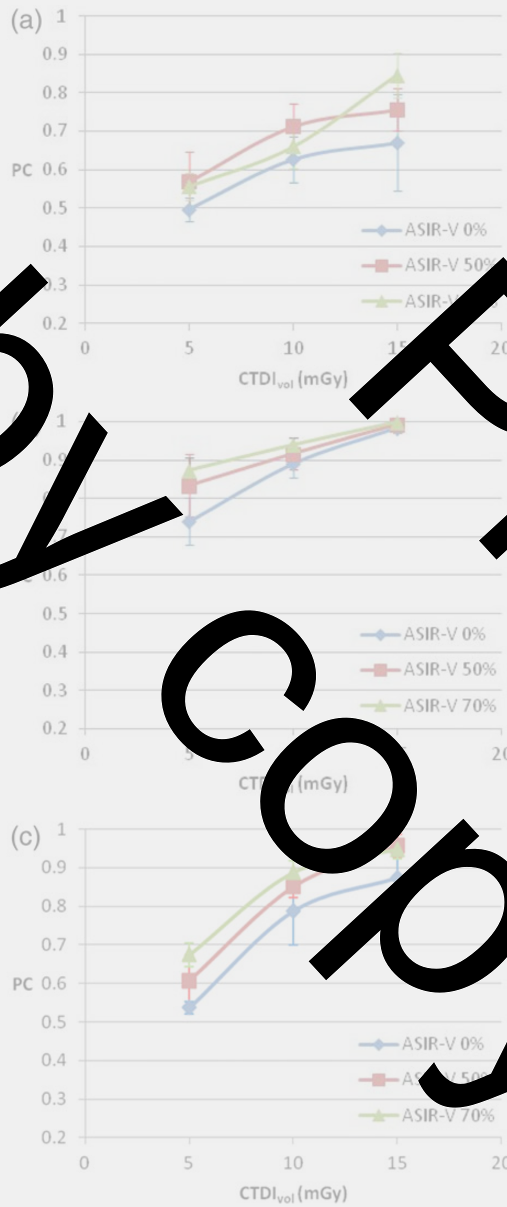

Figure 1. Results of the human observers who performed a 4-AFC test: (**a**) 6 mm/10 HU, (**b**) 6 mm/20 HU and (**c**) 8 mm/10 HU.

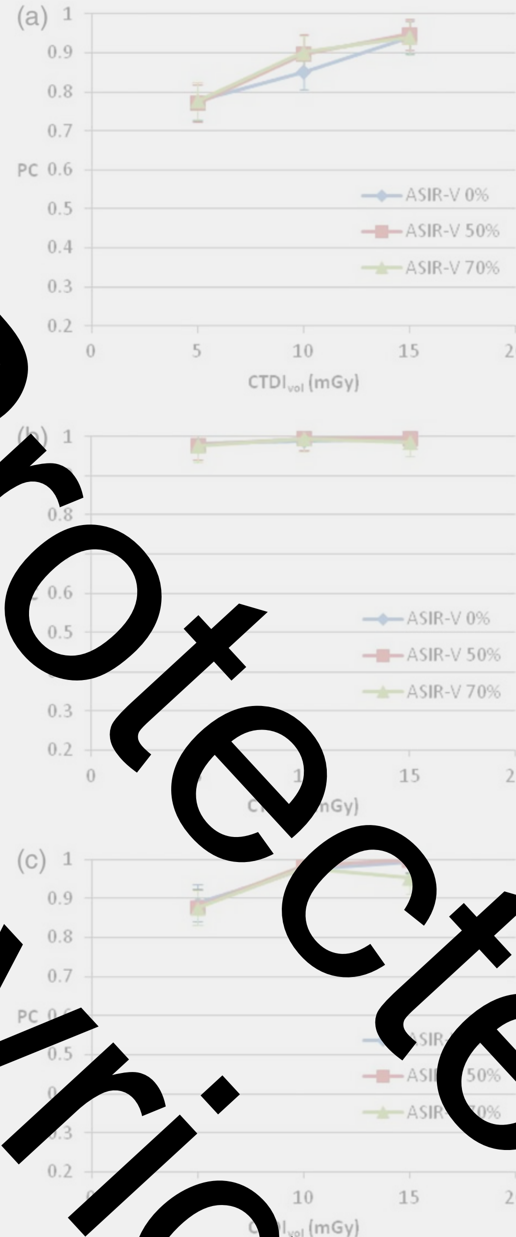

Figure 2. Results of the CHO model observer performing without internal noise and DoG channel in PC for category: (**a**) 6 mm/10 HU, (**b**) 6 mm/20 HU and (**c**) 8 mm/10 HU.

for the sphere at 6 mm and 10 and 20 HU (Figure 4a and b). For the other levels, the image quality increased by 10 %, whereas the dose was increased by 50 %.

DISCUSSION

In recent years, the introduction of iterative algorithms has led the way to optimising clinical protocols and possibly decreasing the collective effective dose, especially in CT. The aim of this study was to evaluate the impact of the new ASIR-V algorithm on image quality in order to optimise the patient's exposure using a model observer. The results show that the image quality increased with the level of ASIR-V, but

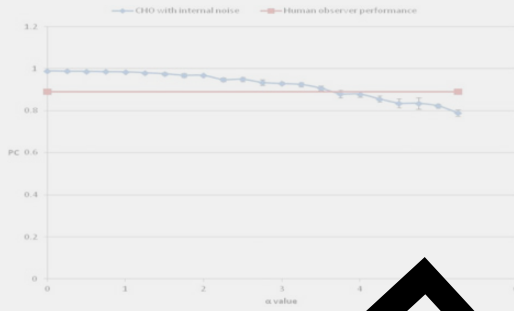

Figure 3. Calibration of the internal noise.

only for the human observer. The human observer seemed more efficient at detecting a signal in a less noisy image, but the results were not statistically significant ($p = 0.26$). The model observer with internal noise function selected allowed to get a better match with human detection performance. Nevertheless, the results could be improved by using more functions since the type of image texture varied in a wide range (new algorithm allowing various strength levels). In spite of this, the authors preferred to use a unique noise function since only one iterative algorithm was used. Thus, the RMSE was just minimised until reaching a plateau, and note that the RMSE was very dependent on an outlier (Figure 4c point at 15 mGy and ASIR-V 70 %). The PC tends to plateau when the dose, the contrast or size increases, and no difference was seen between levels of ASIR-V 50 and 70 %, because this study only focused on the low-contrast detectability and high-contrast performances were not evaluated. To evaluate the interest of the ASIR-V 70 % or higher, the spatial resolution could be evaluated with the target transfer function[19].

One limitation of this study is the use of a simplistic anthropomorphic phantom. Even if it mimics human body attenuations, it does not contain any texture. Furthermore, only two sizes and two contrasts were investigated. Another limitation is that the paradigm used (signal, location and background are known exactly) was simple and therefore different from actual clinical conditions.

CONCLUSION

Evaluating image quality with frequency metrics is far away from the clinical task. To be close to the concerns of radiologists, task-based tools (e.g. CHO model) must be used to objectively evaluate the image quality. The CHO model with DDoG filter used in this study successfully demonstrated its capacity to mimic the human's performance. Thus, the ASIR-V algorithm evaluated with this tool shows that the image quality on the low-contrast detectability stays high even with the small sphere of low contrast. These findings suggest that patient dose could be reduced using the ASIR-V algorithm and the new CT unit evaluated, without decreasing image quality.

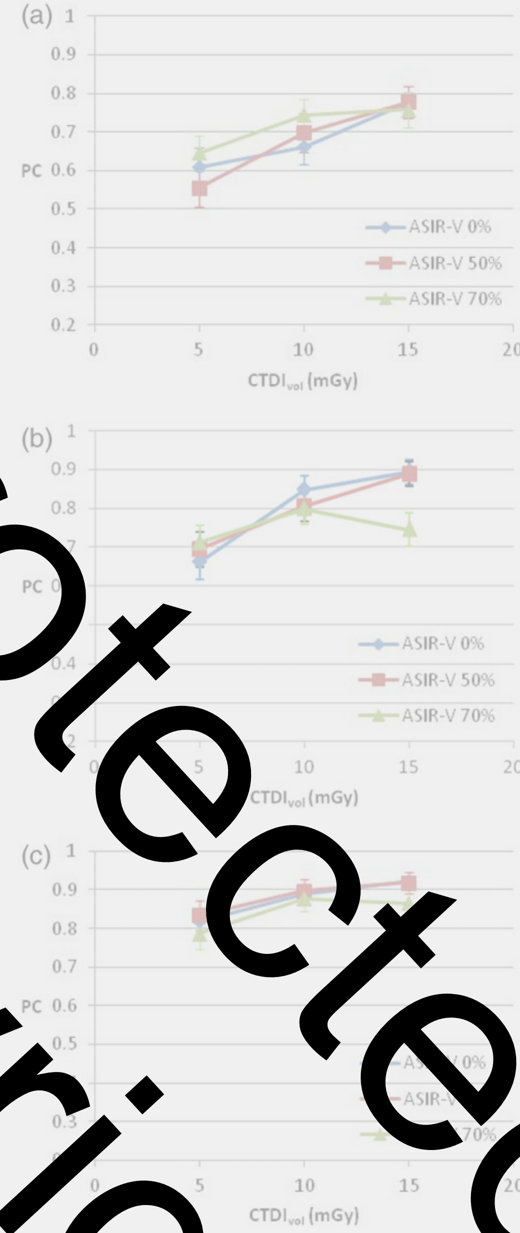

Figure 4. Results of the CHO model observer performance with internal noise and DDoG channel in PC for category: (a) 6 mm/10 HU, (b) 6 mm/20 HU and (c) 8 mm/10 HU.

REFERENCES

1. Samara, E. T., Aroua, A., Bochud, F. O., Ott, B., Theiler, T., Treier, R., Trueb, P. R., Vader, J. -P. and

Verdun, F. R. *Exposure of the Swiss population by medical x-rays: 2008 review.* Health Phys. **102**, 263–270 (2012).

2. Barrett, H. H. and Myers, K. J. *Foundation of Image Science.* John Wiley & Sons (2004).
3. Vaishnav, J. Y., Jung, W. C., Popescu, L. M., Zeng, R. and Myers, K. J. *Objective assessment of image quality and dose reduction in CT iterative reconstruction*. Med. Phys. **41**, 071904 (2014).
4. International Commission on Radiation Units and Measurements. *Receiver operating characteristic analysis in medical imaging*. ICRU Report 79. Journal of the ICRU 8(1) (2008).
5. Metz, C. E. *ROC methodology in radiologic imaging.* Invest. Radiol. **21**, 720–733 (1986).
6. Barrett, H. H., Myers, K. J., Hoeschen, C., Kupinski, M. A. and Little, M. P. *Task-based measures of image quality and their relation to radiation dose and patient risk.* Phys. Med. Biol. **60**, R1–R75 (2015).
7. Leng, S., Yu, L., Zhang, Y., Carter, R., Toledano, A. Y. and McCollough, C. H. *Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain*. Med. Phys. **40**, 0[illegible]08 (2013).
8. Yu, L., Leng, S., Chen, L., Kofler, J. M., Carter, R. E. and McCollough, C. H. *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. **40**, [illegible]41908 (2013).
9. Brunner, C. C., Abboud, S. [illegible], Hoeschen, C. and Kyprianou, I. S. *Signal detection and location-dependent noise in cone-beam computed tomography using the spatial definition of the Hotelling SNR.* Med. Phys. **39**, 3214–3228 (2012).
10. International Electrotechnical Commission. International Standard IEC 60601-2-44. *Medical Electrical Equipment – Part 2-44: Particular Requirements for the Basic Safety and Essential Performance of X-ray Equipment for Computed Tomography*, 3rd edn. [illegible] (20[illegible]9).
11. Lim, K., Kwon, H., Cho, J., Oh, J., Yoon, [illegible], M., Ha, D., Lee, J. and Kang, E. *Initial phantom study comparing image quality in computed tomography using adaptive statistical iterative reconstruction and new adaptive statistical iterative reconstruction v.* J. Comput. Assist. Tomogr. **39**, 443–448 (2015).
12. Singh, S., Kalra, M. K., Hsieh, J., Licato, P. E., Do, S., Pien, H. H. and Blake, M. A. *Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques.* Radiology **257**, 373–383 (2010).
13. Hara, A. K., Paden, R. G., Silva, A. C., Kujak, J. L., Lawder, H. J. and Pavlicek, W. *Iterative reconstruction technique for reducing body radiation dose at CT: feasibility Study.* AJR Am. J. Roentgenol. **193**, 764–771 (2009).
14. Scheffel, H., Stolzmann, P., Schlett, C. L., Engel, L. -C., Major, G. P., Károlyi, M., Do, S., Maurovich-Horvat, P. and Hoffmann, U. *Coronary artery plaques: cardiac CT with model-based and adaptive-statistical iterative reconstruction technique.* Eur. J. Radiol. **81**, e363–e369 (2012).
15. [illegible]no, M., Fuentes, J. M., Singh, S., Hahn, P. F. and Sahani, D. V. *Iterative reconstruction techniques in abdominal CT: technical concepts and clinical implementation*. AJR Am. J. Roentgenol. **205**, W19–W31 (2015).
16. Ott, J. G., [illegible], A., Racine, D., Ryckx, N., Bochud, F. O., Alkadhi, [illegible] and Verdun, F. R. *Patient exposure optimisation through task-based assessment of a new model-based iterative reconstruction technique*. Radiat. Prot. Dosim. DOI: [illegible].1093/rpd/ncw019.
17. Efron, B. and Tbishirani, R. J. *An Introduction to the Bootstrap.* CRC Press (199[illegible]).
18. Castella, [illegible], Eckstein, M. P., Abbey, C. K., Kinkel, K., Verdun, F. R., Saunders, R. [illegible], Samei, E. and Bochud, F. O. *Mass detection in mammograms: influence of signal shape uncertainty on human and model observers.* J. Opt. Soc. Am. A [illegible]5–43[illegible] (2009).
19. Ott, J. G., Becce, F., Monnin, [illegible], Schmidt, [illegible], Bochud, F. O. and Verdun, F. R. *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. **5[illegible]**, [illegible]047–4064 (2014).

Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners

Damien Racine[a)]* and Anaïs Viry*
Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland

Fabio Becce and Sabine Schmidt
Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Rue du Bugnon 46, 1011 Lausanne, Switzerland

Alexandre Ba and François O. Bochud
Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland

Sue Edyvean
Medical Dosimetry Group, Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Didcot OX11, UK

Alexander Schegerer
Department of Medical and Occupational Radiation Protection, Federal Office for Radiation Protection (BfS), Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

Francis R. Verdun
Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland

(Received 30 September 2016; revised 3 February 2017; accepted for publication 7 March 2017; published 13 September 2017)

Purpose: We sought to compare objectively computed tomography (CT) scanner performance for three clinically relevant protocols using a task-based image quality assessment method in order to assess the potential for radiation dose reduction.
Methods: Four CT scanners released between 2003 and 2007 by different manufacturers were compared with four CT scanners released between 2012 and 2014 by the same manufacturers using ideal linear model observers (MO): prewhitening (PW) MO and channelized Hotelling (CHO) MO with Laguerre-Gauss channels for high-contrast spatial resolution and low-contrast detectability (LCD) performance, respectively. High-contrast spatial resolution was assessed using a custom-made phantom that enabled the computation of the target transfer function (TTF) and noise power spectrum (NPS). Low-contrast detectability was assessed using a commercially available anthropomorphic abdominal phantom providing equivalent diameters of 24, 29.6, and 34.6 cm. Three protocols were reviewed: a head (trauma) and an abdominal (urinary stones) protocol were applied to assess high-contrast spatial resolution performance; and another abdominal (focal liver lesions) protocol was applied for LCD. The liver protocol was tested using fixed and modulated tube currents. The PW MO was proposed for assessing high-contrast detectability performance in the various CT scanners.
Results: Compared with older generation CT scanners, three newer systems displayed significant improvements in high-contrast detectability over that of their predecessors. A fourth, newer system had lower performance. The CHO MO was appropriate for assessing LCD performance and revealed that an excellent level of image quality could be obtained with newer scanners at significantly lower dose levels.
Conclusions: This study shows that MO can objectively benchmark CT scanners using a task-based image quality method, thus helping to estimate the potential for further dose reductions offered by the latest systems. Such an approach may be useful for adequately and quantitatively comparing clinically relevant image quality among various scanners. © *2017 The Authors. Medical Physics published by Wiley periodicals, Inc. on behalf of American Association of Physicists in Medicine.* [https://doi.org/10.1002/mp.12224]

Key words: computed tomography, high-contrast spatial resolution, image quality, low-contrast detectability, model observers

1. INTRODUCTION

In most Western countries, the radiation exposure of the population due to computed tomography (CT) examinations has increased steadily for 20 yr.[1] A survey performed in 2006 in the United States showed that the average effective dose due to CT reached 1.5 mSv per capita, per year.[2] The last surveys performed in Switzerland in 2008 and 2013 showed a similar trend, with the average dose per capita from CT increasing from 0.8 mSv to 1.0 mSv within this 5-year period.[3,4]

In this context, the radiation protection requirements in diagnostic radiology (justification of the examination and optimization of the imaging protocol) need to be reinforced. The justification aspect is beyond the scope of this article. The optimization of a CT examination is achieved when image quality enables the clinical question to be answered while keeping patient radiation dose as low as reasonably achievable (ALARA).[5] This goal is, however, difficult to apply in practice. The quality of a CT examination depends on a wide range of parameters such as acquisition time, temporal resolution, and energy resolution when dealing with kV optimization or spectral CT imaging, and other factors. Thus, the actual determination of the clinical performance of a CT scanner is quite complex, and the clinical question needs to be clarified to enable a standard for image quality level to be set. Task-oriented image quality criteria can then be used as surrogates for the assessment of actual clinical image quality.[7–10] These will necessarily be simple in comparison to the clinical situations, but will make it possible, for example, to predict the ability to detect simple structures of high and low contrast within homogeneous backgrounds.[9,10,11] This represents the most basic task that can be considered a surrogate for measuring clinical image quality.

To add complexity, making the task more realistic, one could then not only consider the detection but also the determination of the correct position of the detected structure. Then, the performance with which these tasks are performed could be assessed over more realistic structured backgrounds that mimic the actual anatomy. To go a step further, one could also check if the sizes and contrasts measured on the images correspond to the actual values. This strategy is still far from actual clinical image quality assessment, but used in the optimization of clinical protocols.

The aim of this study was to propose a way to objectively compare CT scanner performance using the simplest task-based image quality assessment: detection. This method was used in particular to evaluate the impact of technological developments on the potential for radiation dose reduction. In addition, we also wanted to investigate if major differences in performance existed among different manufacturers in the limited image quality criteria chosen. We compared the outcomes of four CT scanners released by the four major manufacturers from 2003 to 2007 with the outcomes of newer systems introduced from 2012 to 2014, using ideal model observers (MO) on three clinically relevant protocols.

2. MATERIALS AND METHODS

2.A. CT scanners and clinical protocols

This study was conducted using eight different CT scanners: two per major manufacturer including models released between 2003 and 2007 (referred as "older"), and models released between 2012 and 2014 (referred as "newer"). These eight CT scanners were, listed as "older"/"newer": LightSpeed VCT/Revolution CT (GE Healthcare, Milwaukee, WI, USA), Brillance 40/Ingenuity Core 128 (Philips Medical Systems, Best, the Netherlands), Somatom Sensation 64/Force (Siemens Healthcare, Forchheim, Germany), and Activion 16/Aquilion Prime (Toshiba Medical Systems, Tokyo, Japan). Basic characteristics and the image reconstruction methods used for this study are summarized in Table I.

For all CT systems, the displayed weighted computed tomographic dose index ($CTDI_w$) data were verified by measuring the normalized weighted computed tomographic dose index ($nCTDI_w$) using a 32-cm diameter CTDI test object and a 10-cm long CT pencil ionization chamber connected to an electrometer (model 1035-10.3 CTDI chamber and MDH model 1015 electrometer, Radcal, Monrovia, CA, USA), calibrated in RQR9 and RQA9 beams according to IEC 61267 and traceable to the Swiss Federal Office of Metrology.[12] The volume CT dose index ($CTDI_{vol}$) is defined as the $CTDI_w$ divided by the helical pitch factor; the values used in this study were taken directly from the displayed ones.

The image acquisition protocols used to compare the performance of the CT units were proposed by a panel of four senior radiologists working in three different University Hospitals in Switzerland. Among a large number of clinically relevant protocols we focused on three: two requiring a relatively high level of spatial resolution for the detection of high-contrast structures in the head and abdomen, and one requiring a high level of low-contrast resolution in the abdominal region. When dealing with the assessment of low-contrast resolution performance in the abdominal region, two approaches were chosen: one using fixed dose levels (5, 10, and 15 mGy), with the 15 mGy dose level corresponding to the Swiss abdominal CT Diagnostic Reference Level (DRL) using one phantom size and the other using the tube current modulation option (with the local settings used for the clinical indication of the acquisitions) using three phantom sizes.[13] The details of the acquisition parameters used for each protocol are given in Table I. For technical reasons, the acquisition parameters used were not exactly the same.

2.B. Phantoms and data acquisitions

2.B.1. *High-contrast performance*

The assessment of high-contrast performance for head and abdominal protocols was made using a dedicated custom-made phantom containing cylindrical rods of different contrast materials (Teflon or polytetrafluoroethylene [PTFE], polyethylene, and polymethyl methacrylate [PMMA])[14] (Fig. 1). The edge of this internal cylinder, at different z-axis positions, is used as the interface for the high-contrast numerical evaluation. The external diameter of this phantom is 250 mm; the high-contrast internal cylinder diameter is 100 mm.

2.B.2. *Low-contrast performance*

A modified anthropomorphic abdominal phantom (QRM 401, QRM, Moehrendorf, Germany) (Fig. 2) was used to investigate the low-contrast resolution performance of the CT units. It is made of calibrated tissue-equivalent materials.

TABLE I. Acquisition parameters for each clinical protocol.

Manufacturer	GE Healthcare		Philips		Siemens		Toshiba	
CT Scanner	Light speed	Revolution	Brilliance	Ingenuity core	Sensation	Force	Activion	Aquillion
Release year	2005	2014	2006	2011	2003	2012	2007	2012
Rows	64	256	64	256	64	192	16	80
Algorithms	FBP	ASIR-V	FBP	Idose	FBP	Admire	FBP	AIDR 3D
Voltage	120 kV	120 kV	120 kV	120 kV	120 kV	120 kV	120 kV	120 kV
Acquisition mode	Helical	Helical	Helical	Helical	Helical	Helical	Helical	Helical
Slice Thickness	2.5 mm	2.5 mm	2.5 mm	2.5 mm	3 mm	3 mm	3 mm	3 mm
Head HR: Trauma								
FOV	25 cm							
$CTDI_{vol}$	15 mGy							
Kernel	Bone Plus	Bone Plus	D	D	H70h	Hr64h	FC80	FC30
Pitch	0.984	0.969	1	1	1	1	0.825	0.825
Abdomen HR: Search for urinary stones								
FOV	25 cm							
$CTDI_{vol}$	15 mGy							
Kernel	Standard	Standard	B	B	B31s	Bf40s	FC18	FC18
Pitch	1.375	1.375	1.295	1.295	1.4	1.4	1.475	1.475
Abdomen LCD: Detection of focal liver lesions								
FOV	32 cm							
$CTDI_{vol}$	5-10-15 mGy							
Kernel	Soft	Soft	B	B	B31s	Bf40s	FC18	FC18
Pitch	0.984	0.969	1	1	1	1	0.825	0.825
FOV	32-36-42 cm							
$CTDI_{vol}$	Automatic tube current modulation							
Kernel	Soft	Soft	B	B	B31s	Bf40s	[illegible]	FC18
Pitch	1.375	1.375	1.295	1.295	1.4	1.4	1.475	1.475

The body of the phantom (equivalent diameter of 24 cm) contains muscle, liver, spleen, and bone (vertebrae) tissue equivalents. A module can be inserted into the phantom body that includes spheres of different sizes: 8, 6, 5, 4, and 3 mm; each size having a contrast 20 HU relative to the background at 120 kV. For practical reasons, only three spheres (5, 6, and 8 mm) were used.[15] Two additional annuli (increasing the phantom's effective diameter to 29.6 cm and 34.6 cm, respectively) were added to simulate a range of body habitus (from an approximate patient weight of 50 kg for the equivalent diameter of 24 cm to 75 kg and 100 kg for the equivalent diameter of 29.6 cm and 34.6 cm).

Ten successive scans of the phantom fixed in place were performed to obtain 40 regions of interest (ROIs) with the spheres, and 40 ROIs without any target. This phantom was scanned using two protocols. First, the small phantom was

FIG. 1. Photo and sagittal view of the TTF phantom filled with water with its different contrast rods: PTFE, polyethylene, and PMMA.

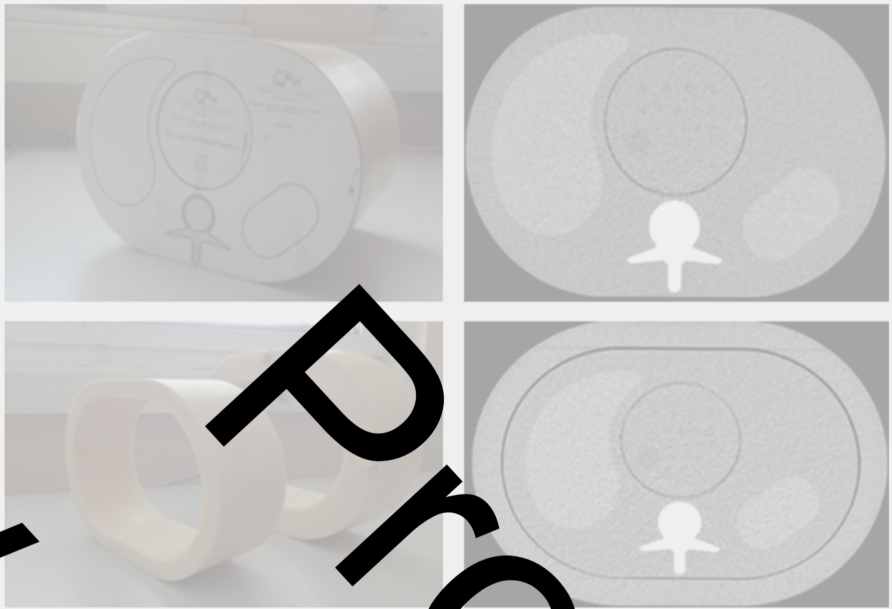

Fig. 2. Pictures of the QRM 3D LCD phantom and additional annuli (left) and example of images (right) provided by the QRM abdominal LCD phantom (equivalent diameters of 24 and 28 cm).

scanned at three dose levels to assess the baseline values of the CT scanners. Then, images were again acquired, this time with the two additional rings installed, to investigate the effect of body habitus on low-contrast detectability (LCD).

2.C. Image quality metrics

2.C.1. High-contrast performances

Spatial resolution: The parameter usually used to assess spatial resolution when dealing with CT images is the Modulation Transfer Function (MTF). However, iterative reconstruction (IR) algorithms are known to be highly nonlinear and therefore might introduce a dependency of the image contrast over the spatial resolution. Boone[16] and Richard et al.[17] proposed target transfer function (TTF) metrics to overcome this problem by characterizing spatial resolution taking into account contrast properties. MTF and TTF are similar metrics, except that TTF may be applied on different contrast levels. In this study we took a similar approach. Using the rod phantom, TTF was calculated for each rod from the radial mean of the edge spread function (ESF) profiles. The ESF's raw data were fitted and analytically differentiated to provide line spread functions (LSFs). Finally, performing a Fourier Transform on the LSFs gave the TTFs, which were normalized to 1 at the zero frequency. More details about the methods can be found in Ott et al.[14]

Noise power spectrum: The rod phantom also allows the assessment of the noise power spectrum (NPS). ROIs of 100 × 100 pixels, which were located in the center of a homogeneous region from 10 images, were used to calculate the NPS. The 2D NPS was computed using the following equation:

$$NPS_{2D}(f_x,f_y) = \frac{\Delta_x \Delta_y}{L_x L_y} \frac{1}{N_{ROI}} \sum_{i=1}^{N_{ROI}} \left| FT_{2D}\{ROI_i(x,y) - \overline{ROI_i}\} \right|^2 \quad (1)$$

Where Δ_x, Δ_y are the pixel sizes in the x and y dimensions, L_x, L_y are ROI sizes in the two directions ($L_x = L_y = 100$ pixels), N_{ROI} is the number of ROIs ($N_{ROI} = 10$), and $\overline{ROI_i}$ is the mean pixel value for the ith ROI. The 2D NPS was then radially averaged to provide the 1D NPS_{1D} according to the methodology presented in ICRU 87.[18]

Prewhitening model observer: To perform a task-based image quality assessment of high-contrast structures, the detectability index (d′) of different diameters structures having a nominal contrast of 1080 HU at 120 kVp (PTFE/water), 120 HU (PMMA/water), and −80 HU (polyethylene/water) was computed using the prewhitening mathematical model observer PW,[19] and to reduce inconsistencies due to the use of iterative reconstruction, the MTF function, that should be used in the model was replaced by the TTF function (see Eq. 2)[19]:

$$d' = \sqrt{2\pi}|\Delta HU| \sqrt{\int_0^{f_{Ny}} \frac{S^2(f)TTF^2(f)f}{NPS_{1D}(f)f} df} \quad (2)$$

where f is the frequency, f_{Ny} is the Nyquist frequency of the image, $|\Delta HU|$ is the absolute contrast difference between the signal and the background and S(f) is the Fourier transform of the input signal, $S(f) = \frac{R}{f} J_1(2\pi Rf)$ with J_1 a Bessel function of the first kind. In our study, the rod phantom was used to provide the estimation of TTF and NPS, which are needed for PW MO, but not for the direct measurement of the small size disks' high-contrast detectability. However, the this limitation, we simulated a virtual disk with a radius R varying from 0.5 mm to 2.5 mm.

It is of note that scatter reduces not only image contrast but also the amplitude of the TTF. It was decided to take into account the scatter effect by using the measured contrasts rather than the nominal ones. Thus, TTFs were fitted to avoid the effect of scatter (spatial resolution drop in the low frequency range) as presented in reference.[14]

The uncertainty of the PW outcome was assessed by varying randomly the contrast and TTF values in the range of their standard deviations, considering a Gaussian distribution, measured on 30 images. The NPS parameter was not considered due to the fact that its uncertainty is negligible compared to that of the contrast and TTF parameters.

2.C.2. Low-contrast performance

Channelized Hotelling observer: LCD was evaluated in the image domain using a channelized Hotelling observer (CHO), with Laguerre-Gauss (LG) channels. This model is an estimation of the Hotelling observer, which itself is the ideal linear MO. The use of LG channels is appropriate in this case because they are known to maximize task performance.[20,21] The computation was made up to the tenth order of the LG polynomials and for one orientation only (due to the circular symmetry of the structure to be detected), resulting in a total of 10 channels. The pth channel (u_p) is obtained by multiplying the Laguerre polynomial at the pth order by a Gaussian function:

$$\boldsymbol{u}_p(r|a_u) = \frac{\sqrt{2}}{a_u}\exp\left(\frac{-\pi r^2}{a_u^2}\right)L_p\left(\frac{2\pi r^2}{a_u^2}\right) \tag{3}$$

where L_p is a Laguerre polynomial, r is a two-dimensional spatial coordinate, and a_u is the width of the Gaussian function (taken to be = 9 in the present study).

Laguerre polynomials are defined by:

$$L_p(x) = \sum_{k=0}^{p}(-1)^k\binom{p}{k}\frac{x^k}{k!} \tag{4}$$

The image is passed through the 10 LG channels. The channel output is a scalar $\mathbf{v}_i$ obtained by the dot product between the channel u_p and the image g:

$$v_i = U^T g \tag{5}$$

Where U represents the matrix of the channels, each column is one of the 10 channels:

$$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots \boldsymbol{u}_{10}] \tag{6}$$

The CHO is then computed from the template w_{LG}:

$$w_{LG} = \left(K_{v/n}\right)^{-1}(\langle v_s\rangle - \langle v_n\rangle) \tag{7}$$

where $(K_{v/n})$ is the covariance matrix calculated from 100 signal absent images as perceived through the channels (channelized images).$\langle v_s\rangle$ represents the mean of 40 channelized signal images and $\langle v_n\rangle$ the mean of 100 channelized absent signal images.

The decision variable λ_{LG} of the CHO model is obtained by combining the template w_{LG} and the channelized image v_i:

$$\lambda_{i,LG} = w_{LG}^T v_i \tag{8}$$

In the end, the MO was tested with the same set of images as with the training set although this could overestimate its performance.[22] For each category (lesion and phantom size as well as dose levels), a receiver operating characteristic (ROC) curve was calculated with 50 threshold levels.[23] To summarize the information, an area under the ROC curve (AUC) was calculated using the trapezoidal method. The average and standard deviation of the model observers were estimated by performing a bootstrap method.[24] In practice, 500 ROC experiments were performed for each category.

For each dose level and phantom size, we used an image quality metric called "AUC_w" which combines the AUCs for lesions of different sizes. This metric is computed thus:

$$AUC_w = \frac{\sum_{i\in(8;6;5)}\frac{AUC_{\text{lesion(i)}}}{i}}{\sum_{i\in(8;6;5)}\frac{AUC_{\text{lesion(i,max)}}}{i}} \tag{9}$$

where i represents the lesion sizes: 8, 6, or 5 mm, $AUC_{\text{lesion(i)}}$ represents the outcome of model observer for each lesion size. With such a definition, $AUC_{\text{lesion(I,max)}}$ corresponds to the value of this metric when the performance is maximal for each lesion size ($AUC_{\text{lesion(I,max)}} = 1.0$).

3. RESULTS

To ensure the impartiality of this work, the results are reported in an anonymous manner consistently throughout the manuscript. A capital letter (A, B, C, and D) was assigned to each manufacturer and the lower case letter "a" and "b" was added for respectively "newer" and "older" CT units. Differences between the displayed and measured $CTDI_{vol}$ were within 15%, in conformity with Swiss legal requirements (limit of ± 20%).

3.A. Image quality for high-contrast structures

3.A.1. High-contrast detection for the head protocol

For the detection of high-contrast structures, a d′ was calculated for different contrast values for each CT using a head protocol. As expected, the d′ increased with the diameter and the nominal contrast of structures to be detected (Figs. 3(a)–3(c).

Comparison of performance of new and old scanner models from each manufacturer: Figure 3(a) shows that for manufacturers A, C, and D, there was a noticeable improvement of the detectability when switching from the older CT scanners to newer ones while a slight reduction was observed for manufacturer B. The largest improvement was observed for manufacturer C (283% for lesions from 3–5 mm), whereas moderate improvements were found for manufacturers A and D (18% and 37%, respectively).

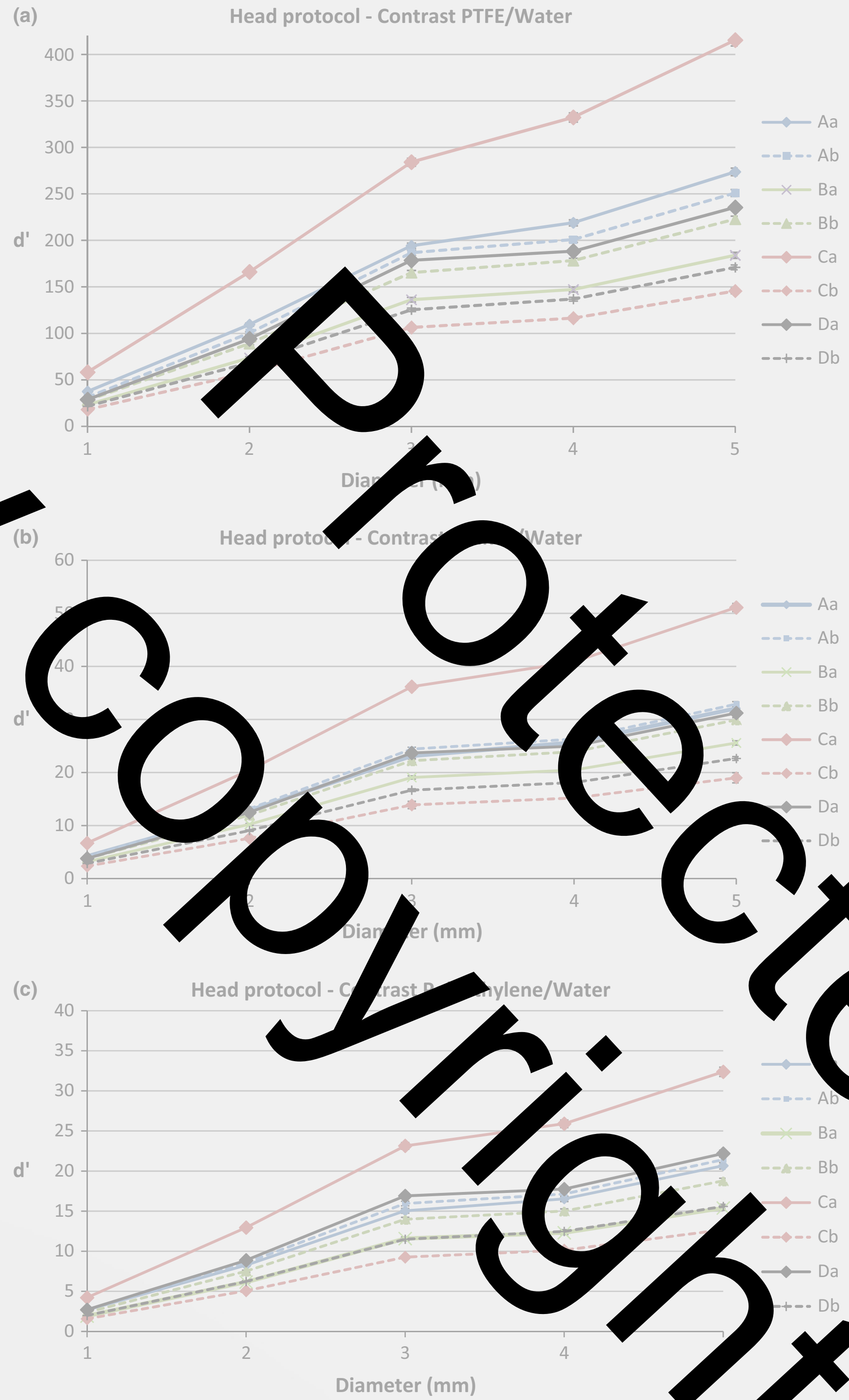

FIG. 3. (a) Detectability index (d′) calculated with the PW model for a nominal contrast of 1080 HU. The horizontal axis represents the diameter of structures. Dotted lines represent older CTs and solid lines represent newer CTs. (b) Detectability index (d′) calculated with the PW model for a nominal contrast of 120 HU. The horizontal axis represents the diameter of structures. Dotted lines represent older CTs and solid lines represent newer CTs. (c) Detectability index (d′) calculated with the PW model for a nominal contrast of −80 HU. The horizontal axis represents the diameter of structures. Dotted lines represent older CTs and solid lines represent newer CTs.

In Figs. 3(b) and 3(c), similar behavior was observed for manufacturers B, C, and D. For manufacturer A, no major difference appeared between the older and newer CT units.

Differences between manufacturers (newer CT models). For the three contrast levels tested with the newer CT models presented in Figs. 3(a)–3(c), the d′ reached the highest value for manufacturer C, and manufacturers A and D provided better results than manufacturer B.

High-contrast detection for the abdomen protocol: The same methodology was applied to assess the detectability of high-contrast structures for the abdominal protocol.

Comparison of performance of new and old scanner models from each manufacturer: In Fig. 4(a), for the highest contrast level, detectability improved when switching from older to newer CTs for manufacturers A, C, and D. A major improvement was noted for manufacturer C (86%) and moderate improvement was noted for manufacturer A and D ([illegible] and 40%, respectively). For manufacturer B, a trend similar to the one identified in the head protocol was observed (Fig. 3[illegible]).

In Figs. 4(b) and 4(c), similar behaviors were observed for manufacturers B, C, and D. For manufacturer A, smaller differences appeared between the older and newer CT units for materials PMMA and polyethylene than for PTFE.

Differences among manufacturers (newer CT models): For each contrast level and each structure diameter, the results of the comparison of newer CT models was very similar to the results for the head protocol. Manufacturer C reached the highest performance. Manufacturers A and D provided better results than manufacturer B.

3.B. Image quality for low-contrast detectability

3.B.1. Abdomen low-contrast detection — $CTDI_{vol}$ variation

Imaging the small abdomen phantom with a $CTDI_{vol}$ of 15 mGy (Fig. 5) showed no major differences among the various scanners. 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 as an uncertainty of 0.003 was set for these measurements ($P < 0.05$). At the lowest $CTDI_{vol}$, we investigated (5 mGy), all newer scanners provided better results than the older ones except for scanner "Aa".

To investigate the robustness of the method used, the measurements were repeated five times on the same scanner "Da" using the small abdomen phantom with a $CTDI_{vol}$ of 5 mGy (Fig. 5) under "positioning uncertainties," and demonstrate that comparable results could be obtained when repositioning the phantom several times.

To investigate how the method would vary when characterizing various scanners of the same type, the methodology was applied on five different "Da" scanners (Fig. 5), and represented as "CT machine uncertainties." Comparable results were found with different machines of the same type.

3.B.2. Abdomen low-contrast detection – Phantom size variation

Using automatic tube current modulation and the small abdominal phantom, Fig. 6 shows that 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 noticeable different $CTDI_{vol}$ values (almost 300%).

When using the medium abdominal phantom a significant drop in image quality is observed for three scanners ("Db", "Bb", and "Ba"). For the other scanners, comparable image quality is preserved but again within a large range of $CTDI_{vol}$ values (Fig. 7).

Finally, when using the largest anthropomorphic abdominal phantom, large differences in behaviors were observed (Fig. 8).

Comparison of performance of new and old scanner models from each manufacturer: For all manufacturers but one, major improvements were demonstrated with the newer models. It was possible to reach similar image quality levels at significantly lower $CTDI_{vol}$ levels. For manufacturer A, image quality level was slightly decreased but the patient exposure was reduced by [illegible]0%. For manufacturer C, a noticeable improvement in image quality was obtained at less than half the dose from the older model. For manufacturer D, a major improvement of image quality was obtained with a lower dose reduction than manufacturers A and C (30%). The only manufacturer where no major improvement was noted was manufacturer B, where similar image quality level was obtained at a slightly higher $CTDI_{vol}$ value.

Differences between manufacturers (newer CT models): When using the largest size of the phantom (simulating a patient [illegible]100 kg), all newer CT scanners reached a [illegible] level of image quality ($AUC_w >$ [illegible]850). Nevertheless, this level of image quality was reached with $CTDI_{vol}$ differences within a range of 300[illegible].

4. DISCUSSION

A full characterization of the scanner units would require the assessment of a large number of parameters. Among these parameters one could mention: the acquisition time, the standard high- and low-contrast resolutions, the temporal resolution, and the energy resolution when dealing with kV optimization or dual energy imaging.

We chose to use simple task-based image quality assessment methodologies that do not include the whole range of potential performance of the scanners. We assessed the performance regarding image quality of high- and low-contrast structures

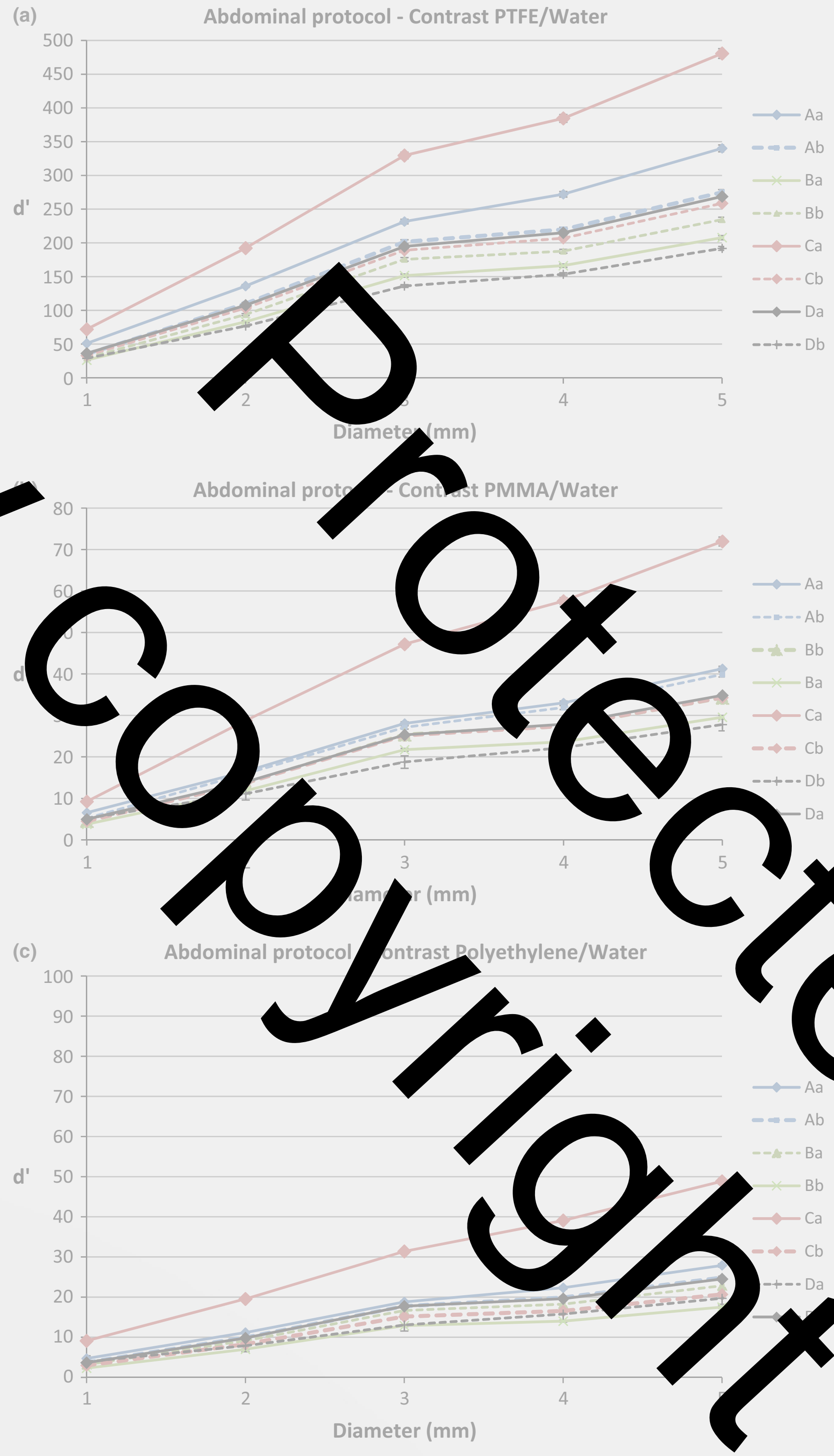

FIG. 4. (a) Detectability index (d′) calculated with the PW model for a nominal contrast of 1080 HU. The horizontal axis represents the diameter of structures. (b) Detectability index (d′) calculated with the PW model for a nominal contrast of 120 HU. The horizontal axis represents the diameter of structures. (c) Detectability index (d′) calculated with the PW model for a nominal contrast of −80 HU. The horizontal axis represents the diameter of structures.

using ideal MO for eight CT systems and three clinical protocols. This benchmark provided a large panel of image quality levels for older (2003–2007) and newer CTs (2011–2014).

The first aim of the study was to investigate if technological improvements over time could be shown using our limited set of image quality criteria. For the high contrast

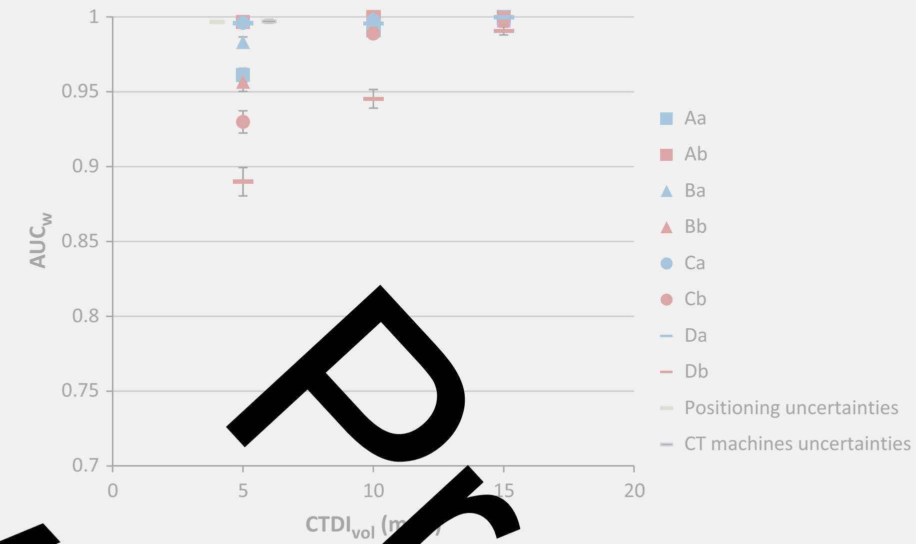

FIG. 5. AUC_w to detect a sphere in soft tissue equivalent at three $CTDI_{vol}$ values using the smallest anthropomorphic abdominal phantom with fixed tube current. The vertical axis is the AUC_w for each of the eight CTs. The errors bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, and the blue color represents the newer ones. AUC_w = weighted area under the curve $CTDI_{vol}$ = volume CT dose index.

FIG. 6. AUC_w to detect a sphere in soft tissue equivalent using the smallest anthropomorphic abdominal phantom at different $CTDI_{vol}$ levels with automatic tube current modulation for the eight different CT scanners. The vertical axis represents the AUC_w outcome. The error bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, whereas the blue color represents the newer ones. AUC_w = weighted area under the curve, $CTDI_{vol}$ = volume CT dose index.

detectability, scanners could be discriminated using the PW MO, and performance improvement was noted for manufacturers A, C, and D. The d' values were systematically very high, indicating that the detection of a structure > 2 mm in diameter with such a nominal contrast value was trivial. For better discrimination one could add complexity to the task, for example: the estimation of shape, size, and contrast. Concerning the low-contrast resolution, performance improvements were observed also for three (manufacturers A, C, and D) out of four manufacturers with a drastic dose reduction to reach similar high image quality levels.

The second aim of our study was to investigate if major differences in performance existed between newer CT scanners of various manufacturers. For the limited criteria chosen in this study, manufacturer [illegible] reached the highest performance with the chosen reconstruction kernels. However, our measurements have two limitations: the first one deals with the choice of the reconstruction kernel that could not be the same for all manufacturers.[25]

To investigate if some kernels used in this study would give advantage to a particular range of target sizes, we computed the d' values for structures 1–5 mm in diameter. A singular increase in d' with the object diameter was noted for all manufacturers. In addition, the clinical parameters used for the high-contrast resolution would not necessarily represent the maximum theoretical capability of the system. The

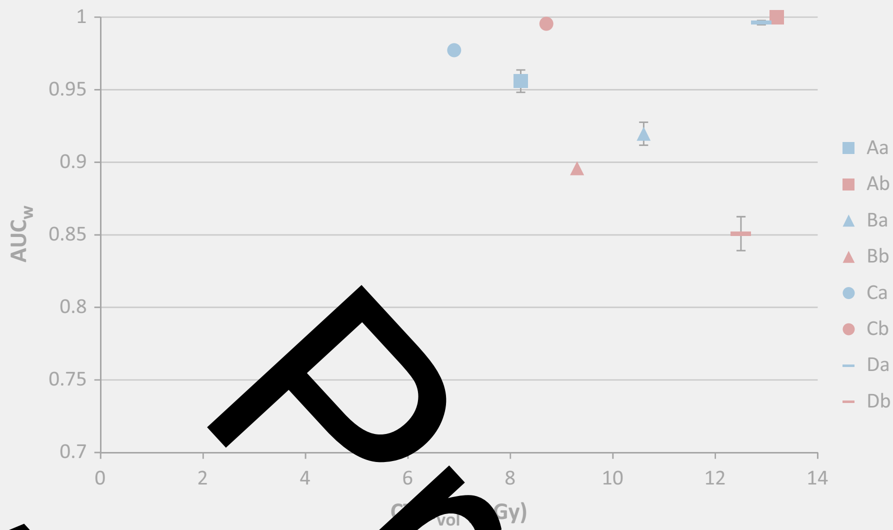

FIG. 7. AUC$_w$ to detect a sphere in soft tissue equivalent using the medium anthropomorphic abdominal phantom at different CTDI$_{vol}$ levels using automatic tube current modulation for the eight different CT scanners. The vertical axis represents the AUC$_w$ outcome. The errors bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, whereas the blue color represents the newer ones. AUC$_w$ = weighted area under the curve, CTDI$_{vol}$ = volume CT dose index.

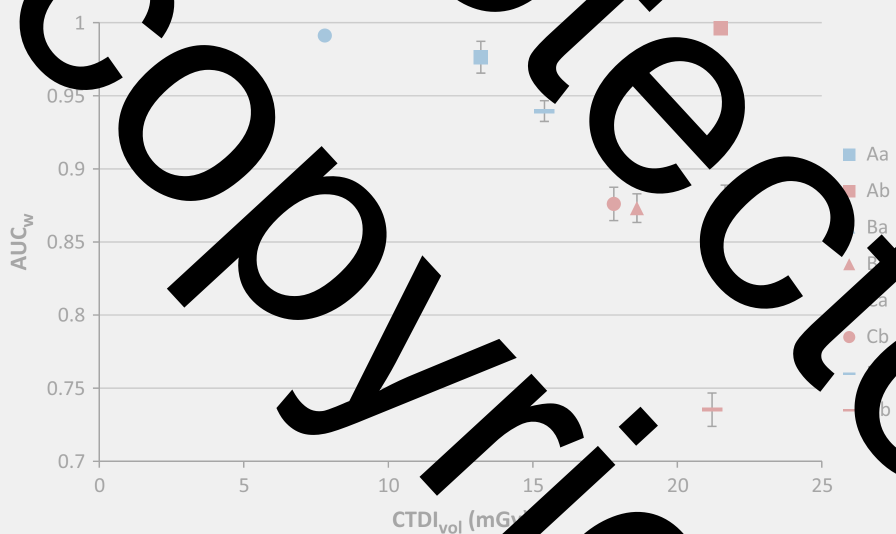

FIG. 8. AUC$_w$ to detect a sphere in liver tissue equivalent using the large phantom at different CTDI$_{vol}$ levels using automatic tube current modulation for the eight different CT scanners. The vertical axis represents the AUC$_w$ outcome. The errors bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, whereas the blue color represents the newer ones. AUC$_w$ = weighted area under the curve, CTDI$_{vol}$ = volume CT dose index.

second limitation deals with the Slice Sensitivity Profile, a parameter that was not considered in our methodology. Nevertheless, comparable reconstruction slice thicknesses (within the range 2.5–3.0 mm) were used, so no major influence of this parameter is expected in our results. Moreover, as the 2D NPS was not isotropic in the phantom, using symmetric channels could impact the performance of the model; but this impact would be minor.

Dealing with the protocol using fixed CTDI$_{vol}$ values, we proposed first to use a range of dose levels. The highest value, 15 mGy, corresponds to our DRL for a "standard" patient of 75 kg. At 15 mGy, no difference between CTs appeared when we used the small abdominal phantom. However, differences were detected at lower dose levels. Thus, such a phantom should be used over a lower dose range (e.g., 2–10 mGy) or replaced by a larger version as shown in Figs. 7 and 8 to provide useful results. Finally, with a larger phantom size, the use of a fixed tube current might be a limitation that introduces weaknesses of the image quality which are not relevant for clinical applications where tube current modulation is generally used.

The use of the protocol with tube current modulation on various phantom sizes is certainly more realistic in the framework of patient dose optimization. When comparing the products of one manufacturer, the results were generally straightforward. The major difficulty of such an approach is to compare various manufacturers who propose different strategies to manage the balance between image quality and patient exposure. In this study we decided to use the local settings. The indication "search for focal liver lesions" requires particularly high image quality. Using the small phantom, a high level of image quality was reached by all CTs using a large range of $CTDI_{vol}$ values. For the larger phantom, a high level of image quality could not be reached by certain units despite the use of a large range of $CTDI_{vol}$ values.

The outcome of this work clearly demonstrates the weaknesses of the DRL concept; indeed, a similar dose level cannot be reached on different scanners without impairing the LCD. To improve this situation, one could associate the DRL to an image quality criterion such as the LCD estimated in a standardized phantom.

5. CONCLUSION

This study shows that MOs can objectively benchmark CT scanners using a task-based image quality method. Such an approach may be useful for quantitatively comparing the clinically relevant image quality among various scanners to aid in the estimation of the potential dose reduction without missing the detection of critical lesions.

ACKNOWLEDGMENTS

This work was supported by the German Radiation Protection Agency (BfS – "UFO-Plan Vorhabens 3[illegible]07"). The authors thank Prof. Peter Vock and PD Dr Sebastian Schindera, the two other senior radiologists involved in this project for their help when dealing with the choice of the protocols and diagnostic tasks. The authors also thank MSc Pascal Monnin for his advices concerning the estimation of the TTF functions.

CONFLICT OF INTEREST

The authors have no relevant conflicts of interest to disclose.

*D. Racine and A. Viry contributed equally to this work.
[a]Author to whom correspondence should be addressed. Electronic mail: damien.racine@chuv.ch; Telephone: +41 21 314 92 49.

REFERENCES

1. Brenner DJ, Hall EJ. Computed tomography — An increasing source of radiation exposure. *N Engl J Med*. 2007;357:2277–2284.
2. Schauer DA, Linton OW. National council on radiation protection and measurements report shows substantial medical exposure increase. *Radiology*. 2009;253:293–296.
3. Samara ET, Aroua A, Bochud FO, et al. Exposure of the swiss population by medical x-rays: 2008 review. *Health Phys*. 2012;102:263–270.
4. Coultre RL, Bize J, Champendal M, et al. Exposure of the swiss population by radiodiagnostics: 2013 review. *Radiat Prot Dosimetry*. 2015;169:221–224.
5. Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP. Task-based measures of image quality and their relation to radiation dose and patient risk. *Phys Med Biol*. 2015;60:R1.
6. Vennart W. ICRU report 54: medical imaging — The assessment of image quality. *Radiography*. 1996;3:243–244.
7. Racine D, Ott JG, Ba A, Ryckx N, Bochud FO, Verdun FR. Objective task-based assessment of low-contrast detectability in iterative reconstruction. *Radiat Prot Dosimetry*. 2016;169:73–77.
8. Saiprasad G, Filliben J, Peskin A, et al. Evaluation of low-contrast detectability of iterative reconstruction across multiple institutions, CT scanner manufacturers, and radiation exposure levels. *Radiology*. 2015;277:124–133.
9. Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. *Med Phys*. 2014;41:71904.
10. Verdun FR, Racine D, Ott JG, Tapiovaara MJ, et al. Image quality in CT: from physical measurements to model observers. *Phys Medica Eur J Med Phys*. 2015;31:823–843.
11. Schindera ST, Odedra D, Raza SA, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology*. 2013;269:511–518.
12. International Electrotechnical Committee. *Medical diagnostic X-ray equipment - Radiation conditions for use in the determination of characteristics*. IEC Standard IEC # 61267. https://webstore.iec.ch/publication/2[illegible]. Published 2015. Accessed September 27, 2016.
13. Office fédéral de la santé publique OFSP. *Notice R-06-06 :Niveaux de référence diagnostiques en tomodensitométrie*. 2010.
14. Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Verdun FR. 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.
15. Racine D, Ba AH, Ott JG, Bochud FO, Verdun FR. Objective assessment of low contrast detectability in computed tomography with channelized hotelling observer. *Phys Medica Eur J Med Phys*. 2016;32:76–83.
16. Boone JM. Determination of the presampled MTF in computed tomography. *Med Phys*. 2001;28:[illegible]50.
17. Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. *Med Phys*. 2012;39:41[illegible]–4122.
18. International Commission on Radiation Units and Measurements. ICRU report no. 87: radiation dose and image-quality assessment in computed tomography. *J ICRU*. 2012;12:1–149.
19. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *Proc Natl Acad Sci USA*. 1993;90:[illegible]–9765.
20. Park S, Barrett HH, Clarkson E, Kupinski MA, Myers KJ. Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal. *J Opt Soc Am A Opt Image Sci Vis*. 2007;24:B136–B150.
21. Gallas BD, Barrett HH. Validating the use of channels to estimate the ideal linear observer. *J Opt Soc Am A*. 2003;20:1725–1738.
22. Barrett HH, Myers KJ. *Foundations of Image Science*. Wiley-Interscience; 2004.
23. International Commission on Radiation Units and Measurements. Receiver operating characteristic analysis in medical imaging. In: *ICRU Report n 79*, Vol 79. Bethesda, MD: International Commission on Radiation Units and Measurements; 2008: 79.
24. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: CRC Press; 1994.
25. Solomon JB, Christianson O, Samei E. Quantitative comparison of noise texture across CT scanners from different manufacturers. *Med Phys*. 2012;39:6048–6055.

Radiation Protection Dosimetry (2016), Vol. 169, No. 1–4, pp. 78–83
Advance Access publication 2 March 2016
doi:10.1093/rpd/ncw021

BENCHMARKING OF CT FOR PATIENT EXPOSURE OPTIMISATION

Damien Racine*, Nick Ryckx, Alexandre Ba, Julien G. Ott, François O. Bochud and Francis R. Verdun
Institute of Radiation Physics, CHUV, Lausanne, Switzerland

*Corresponding author: damien.racine@chuv.ch

Patient dose optimisation in computed tomography (CT) should be done using clinically relevant tasks when dealing with image quality assessments. In the present work, low-contrast detectability for an average patient morphology was assessed on 56 CT units, using a model observer applied to images acquired with two specific protocols of an anthropomorphic phantom containing spheres. Images were assessed using the channelised Hotelling observer (CHO) with dense difference of Gaussian channels. The results were computed by performing receiver operating characteristics analysis (ROC) and using the area under the ROC curve (AUC) as a figure of merit. The results showed a small disparity at a volume computed tomography dose index ($CTDI_{vol}$) of 15 mGy depending on the CT units for the chosen image quality criterion. For 8-mm targets, AUCs were 0.999 ± 0.018 at 20 Hounsfield units (HU) and 0.927 ± 0.054 at 10 HU. For 5-mm targets, AUCs were 0.947 ± 0.059 and 0.702 ± 0.068 at 20 and 10 HU, respectively. The robustness of the CHO opens the way for CT protocol benchmarking and optimisation processes.

INTRODUCTION

In 1997, the European Council Directive 97/43/EURATOM[1] required that medical physics experts (MPEs) be involved in the optimisation process of radiological diagnostic procedures. That recommendation has been translated into the Swiss Radiological Protection Ordinance and was officially applied in 2008[2]: 'For nuclear medicine applications and for fluoroscopy-guided interventional radiology and computer tomography, the licence holder must periodically enlist the services of a medical physicist'.

The objective is to involve the medical physicist in the optimisation process of medical devices, especially in CT, which is one of the most commonly used tools in the medical imaging field and which is responsible for the highest dose delivered to the population from diagnostic radiology. Evaluating image quality is crucial when optimising or comparing CT devices, reconstruction algorithms and clinical protocols in order to respect the as low as reasonably achievable (ALARA) principle, especially with the introduction of iterative reconstruction into CT because the standard Fourier metrics are no longer applicable[3, 4]. The aim of this study was to provide an objective tool to compare CT protocols and to develop methods to then use the devices at their full potential, i.e. using the minimum dose while maintaining image quality for diagnosis. This optimisation scheme relied on the use of task-based metrics that approximate the clinical task in order to qualify the image quality and facilitate the cooperation between medical physicists, radiographers and radiologists.

Clinical diagnosis is composed of three tasks: the detection task, the localisation task and the characterisation task. In the present case, the detection task performance was evaluated with a mathematical observer model which is used to substitute the human observer. The use of these tools provides an objective metric to evaluate image quality. The advantage of this approach is that it takes into account the entire imaging chain at the same time as it remains as close as possible to a simplified diagnostic task.

During the authors' visits to the various centres, image acquisitions were performed together with the verification of the indicated volume computed tomography dose index ($CTDI_{vol}$) and the assessment of other parameters such as X-ray beam efficiency and the Hounsfield unit (HU) calibration of water at the available X-ray tube voltages[5]. This contribution will focus on a CT benchmarking based on a model observer. The first aim was the assessment of image quality on several CT units at a given dose level. The second aim was the assessment of image quality while using different acquisition parameters.

MATERIALS AND METHODS

Model observer: channelised Hotelling observer

Anthropomorphic model observers are mathematical models based on the statistical decision theory to estimate the detection performance of human observers. In this study, a linear anthropomorphic model was chosen, namely the channelised Hotelling observer (CHO)[6, 7], which is usually used to evaluate the image quality. As Ott *et al.*[8] demonstrated, the CHO can mimic the human observer performance for a detection task. In this case, the methodology used for the CHO is directly inspired by the methodology used by Ott *et al.*, and extensive details can be found in the chapter on the CHO model observer[8].

Performance measurement of the model observer

All decision variables provide a distribution depending on the presence or absence of a signal. In the receiver operating characteristics (ROC) analysis, if the decision variable calculated using an image containing the signal is above the threshold, then the response is considered true positive. If this variable is calculated with the image containing only noise is above the threshold, then the response is considered false positive. For a given threshold, a true positive fraction (TPF) and a false positive fraction (FPF) are obtained, and the ROC curve is then constructed from pairs of TPF and FPF.

The ROC curve can give a complete description of the performance of an observer. To summarize the performance of an observer, the area under the curve (AUC) can be calculated[9]. It ranges from 0 to 1, with a value of 0.5 meaning that the test is not better than tossing a coin to give a positive or negative answer. AUC was estimated using the trapezoidal method with 100 points.

Description of phantom and CT units

An abdominal anthropomorphic phantom (Figure 1) (QRM, Moehrendorf, Germany) was scanned in 56 CT units (≈20 % of all CT units in the country). This phantom is composed of an external shell intended to closely mimic the X-ray attenuation of an abdomen during a CT examination and of two embeddable modules. The materials used represent three densities and anatomical shapes of the human body:

- Liver (55 HU at 120 kV),
- Spleen (55 HU at 120 kV),
- Vertebra (cortical and cancellous bone).

The uniform phantom background is equivalent to the soft tissue of the abdomen (35 HU at 120 kV).

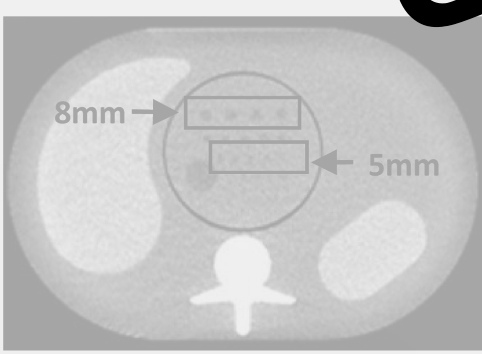

Figure 1. QRM phantom image from an acquisition at a high dose level ($CTDI_{vol}$ 20 mGy). The phantom contains four 8-mm spheres and five 5-mm spheres. The signals represented in this figure have a contrast of 20 HU.

Two modules can be inserted into this external shell: a homogeneous module identical to the background phantom and a module containing low-contrast signals. These signals are spheres of different sizes (8, 6, 5, 4 and 3 mm) with contrasts of −20 and −10 HU at 120 kV with respect to the background.

In 2014, the equipment from four CT manufacturers in Western Switzerland was investigated with the range of the detector coverage represented in Table 1 (in detail 20 devices for Philips, 18 for GE, 12 for Toshiba and 6 for Siemens).

Acquisition protocols

The acquisitions were performed following two protocols:

- One protocol called 'Reference' with the same parameters for all radiological services. Acquisitions were performed at 120 kV, and the tube current was adapted to get indicated $CTDI_{vol}$ values of 15 mGy [the diagnostic reference level (DRL) for the abdominal region in Switzerland]. During the visit, the $CTDI_{vol}$ values were measured on each CT unit using a CTDI phantom of 32 cm in diameter and a 100 mm pencil ionisation chamber as described in the IEC 60601-2-44[10]. The X-ray tube rotation time was 1 s, and the pitch was close to one. The display field of view (DFOV) was 320 × 320 mm for a 512 × 512 matrix size in order to have a constant spatial resolution. Slices were reconstructed with a nominal thickness of [illegible] or 2.0 mm associated to an interval of [illegible] or 2.0 mm (depending on the possibilities offered by the different manufacturers), respectively. Images were reconstructed only with the filtered back-projection (FBP) in the axial plane with a standard convolution kernel used locally for abdominal acquisitions.
- The other protocol called 'Local' where the acquisition parameters were those usually employed by the centres for the detection of focal liver lesions (FBP or iterative reconstruction algorithm was used). The DFOV was then adjusted to 320 × 320 mm for all centres in order to obtain comparable pixel sizes. Since the spread in slice thicknesses

Table 1. Number of CT units included in the study for each detector coverage range.

Collimation (number of rows)	Number of CT units
16	10
32	[illegible]
40	[illegible]
64	36
80	3
128	3

and doses used in the local protocols was large, an alternative metric, called ‘volumetric dose’, was used. The volumetric dose is defined as the product of the dose and the slice thickness, corresponding to the dose length product of the reconstructed slice.

For the two protocols, the phantom was positioned at the isocentre of the CT unit, and the module containing the low-contrast spheres was scanned 10 times successively. Then, the homogeneous module was scanned without changing the position of the abdomen-like shell.

Generating signal-absent and signal-present images

For practical reasons, only two spheres sizes were investigated, namely the 5 and 8 mm (smaller sizes were not visible enough at the dose levels investigated). For each acquisition, 4 centred regions of interest (ROIs) (22 × 22 pixels, 1 pixel = 0.625 mm) per sphere size/contrast combination were extracted from the module containing the sphere, and 8 ROIs per slice were extracted from images of the homogeneous module at the same location (x,y) as the ROI containing the signal. Hereafter, for convenience, ROIs will be called signal-present images if the signal is present in the ROIs and signal-absent images if the signal is absent.

RESULTS

‘Reference’ protocol

Using a CTDI_{vol} of 15 mGy on all systems, the detection performance for a given contrast/diameter combination was globally comparable with a few outliers (Figure 2). Centres 41 and 4[illegible] in Figure 2[illegible] and Centres 43, 45, 46 and 47 in Figure 2c provided a lower-level image quality when compared with the other centres. For 8-mm targets, the averaged AUCs were 0.999 ± 0.018 at 20 HU and 0.927 ± 0.054 at 10 HU. For 5-mm targets, the averaged AUCs were 0.947 ± 0.059 and 0.702 ± 0.068 at 20 and 10 HU, respectively (the error represents one standard deviation). Thus, the relative standard deviation was in the range of 1–10 %, indicating, for that specific image quality criterion and this specific acquisition protocol, that the CT units in Western Switzerland are relatively homogeneous at the quite high dose level investigated. As expected, the detection was greatly enhanced when the diameter increased (5 vs. 8 mm) as well as when the contrast became higher (10 vs. 20 HU).

‘Local’ protocol

When using the local protocol instead of the reference protocol, as expected, it was noticed that the detectability still increased with the contrast and the diameter of the spheres. However, it was also discovered that the mean AUC did slightly change depending on the reconstruction algorithms (AUC was lower in iterative than in FBP), but the mean delivered dose was also much lower for the iterative algorithms (no distinction was made between the level of complexity of the iterative algorithms available) compared with FBP. These measurements clearly highlighted the contribution of the iterative algorithm on dose reduction (almost a factor of two) associated to a slight reduction in low-contrast detectability.

Figure 3 illustrates the detection performance for combinations 8 mm/10 HU and 5 mm/20 HU when switching from the reference to the locally implemented protocols. Varying the parameters used [nominal slice thicknesses (ranging from 0.5 to 7 mm), dose levels, algorithm reconstruction and convolution kernels used for standard abdominal protocols] in each centre slightly increased the variability. Moreover, some centres work with unexpected parameters, which deteriorate image quality and increase the image quality variability. In some cases, the CTDI_{vol} was quite low (3.3 mGy) with very thin slice reconstruction thickness (0.5 mm) or the dose was quite high (12.3 mGy) with the reconstruction of thick slices (7 mm), see three red circles in Figure 3. In the first case, even if the image was reconstructed with an iterative algorithm, the low-contrast detectability could not be recovered. In the second case, the partial volume effect was responsible for the poor score obtained.

DISCUSSION

In this study, an objective task-based image quality assessment was performed using a model observer on an anthropomorphic phantom to assess image quality. The CT units in use in the Western part of Switzerland are relatively recent and homogeneous, but the image quality with the ‘Reference’ protocol was explored at a high dose level. In this case, the discrimination between the CT units is limited, but this benchmark has, however, demonstrated that several units led, for the chosen image quality criterion, to a lower performance. One limit of this ‘Reference’ protocol is that it investigates the image quality at one dose level and only with the FBP algorithm. The benefits of introducing the iterative algorithm, especially the model-based iterative algorithm, were not explored[11]. In future work, the performance of CT units will be investigated in a range of dose and algorithms in order to be able to qualify image quality including high- and low-contrast performances.

When switching to the ‘Local’ protocol, the dose and slice thickness varied in a wide range between the centres and noticeable image quality variations were observed. Based on this metric, it will be possible to optimise and standardise the clinical practice to

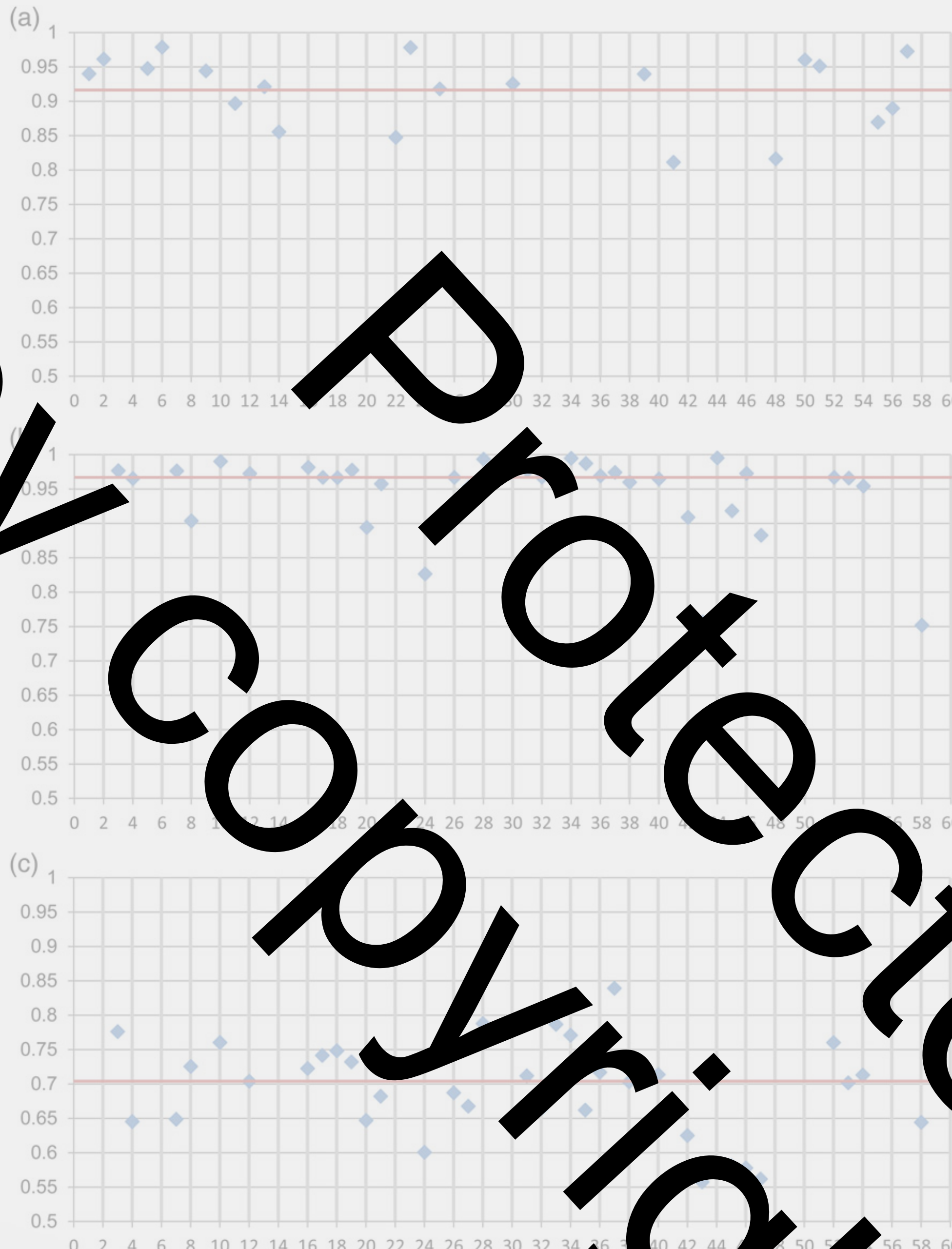

Figure 2. Results of a comparison of the image quality for CT units in Western Switzerland when dealing with the 'Reference' protocol for spheres of 8 mm/10 HU at 2.0 mm slice thickness (**a**), 5 mm/20 HU at 2.5 mm slice thickness (**b**) and 5 mm/10 HU at 2.5 mm slice thickness (**c**). The red line represents the median value.

converge towards a given level of image quality for a particular clinical examination. The main limit of the use of the volumetric dose parameter for that part of the study is the range of slice thickness investigated, since partial volume effects could be very different from one protocol to another. This could be certainly improved by a closer collaboration between the medical physicists and the radiologists when selecting acquisition parameters and the range of reconstructed slice thickness.

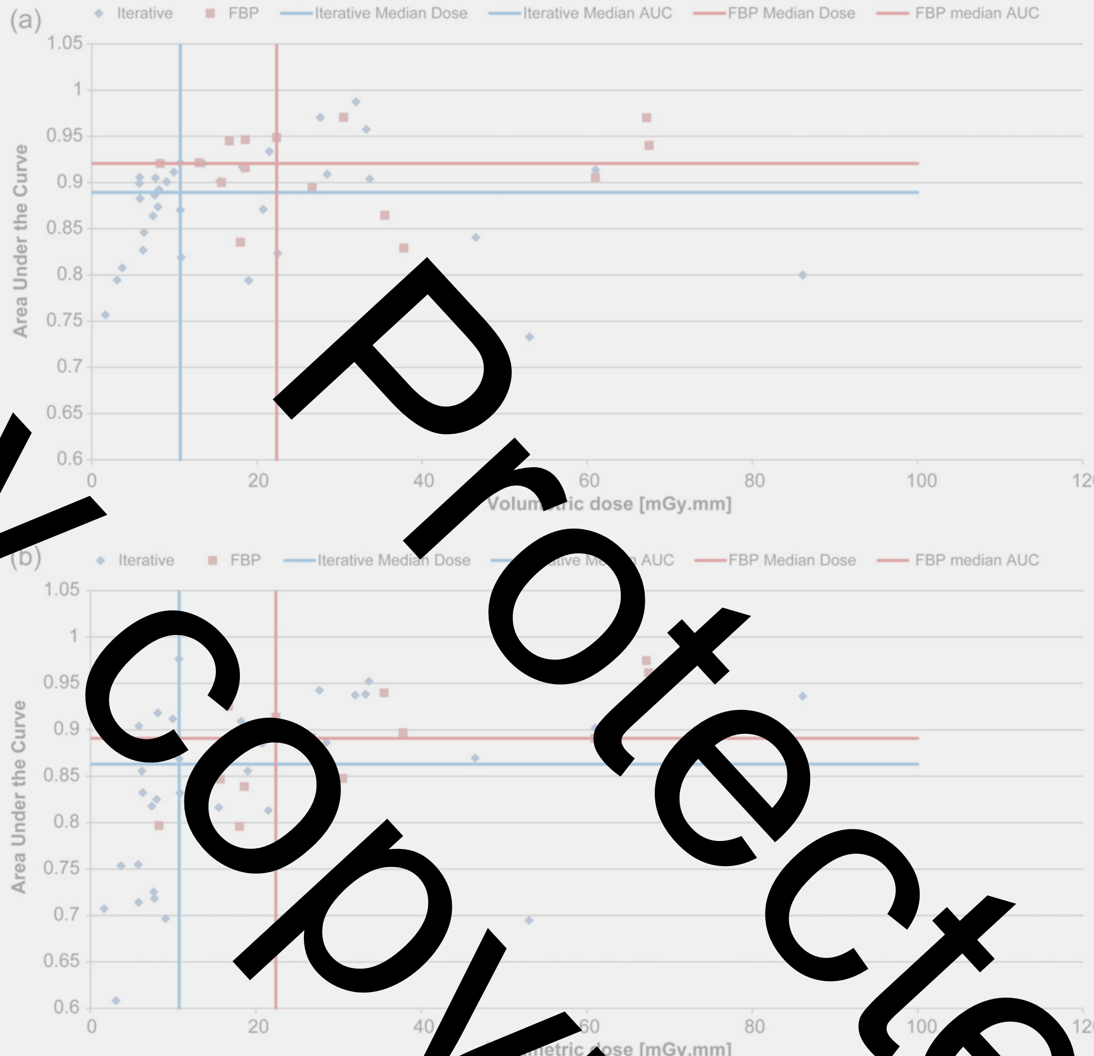

Figure 3. Results of a comparison of [illegible] quality for 5 [illegible] U (**a**) and 8 mm/10 HU (**b**) [illegible] a func[illegible] the volumetric dose ($CTDI_{vol}$ multiplied by the reconstructed [illegible] thickness for taking into account the vari[illegible] slice thi[illegible]nesses produced) regarding the practice in Western Switzerla[illegible] when dealing [illegible] the protocol used for the dete[illegible] of he[illegible] metastasis. The red circles represent the poor image quality o[illegible]ed with unexpected parameters (slice th[illegible]ess inappropriate for th[illegible]ion size).

Thus, the proposed method seems suitable for benchmarking the CT unit as well as the clinical protocols. The method could be improved by adding high Z materials targets, making it possible to optimise the choice of X-ray tube high voltage when dealing with injected protocols.

CONCLUSION

A method for benchmarking CT units has been applied to different CT units in Western Switzerland [illegible] a [illegible]ecific ta[illegible] (i.e. de[illegible]tion of low-contrast spher[illegible]al targ[illegible] appe[illegible]at the image quality was re[illegible]tively [illegible]omoge[illegible]ous. Th[illegible] results also show that the methodolo[illegible] could be [illegible]ed to assess clinically relevant image acquisition [illegible]otoc[illegible]ls [illegible] optimise the process of patient expo[illegible]e. It wa[illegible]served that the use of iterative reco[illegible]uction [illegible]owed a significant dose reduction (almost a f[illegible]r of two) associated, however, with a slight [illegible]uction in low-contrast detectability.

REFERENCES

1. EU. *Council Directive 97/43/Euratom of 30 June 1997 on health protection of individuals against the dangers of ionizing radiation in relation to medical exposure, and repealing Directive 84/466/Euratom*. Official Journal of the European Union **L180**(09/07/1997), 22–25 (1997).
2. The Swiss Federal Council. *Radiological Protection Ordinance (RPO) of 22 June 1994, Classified Compilation 814.501*. The Swiss Federal Council (2008).
3. Barrett, H. H., Myers, K. J., Hoeschen, C., Kupinski, M. A. and Little, M. P. *Task-based measures of image quality and their relation to radiation dose and patient risk*. Phys. Med. Biol. **60**, R1–R75 (2015).
4. Brunner, C. C., Abboud, S. F., Hoeschen, C. and Kyprianou, I. S. *Signal detection and location-dependent noise in cone-beam computed tomography using the spatial definition of the Hotelling SNR*. Med. Phys. **39**, 3214–3228 (2012).
5. Ryckx, N., Gnesin, S., Meuli, R., Elandoy, C. and Verdun, F. R. *Medical physicists' implication in radiological diagnostic procedures: results after 1 y of experience*. Radiat. Prot. Dosim. **164**, 120–125 (2015).
6. Barrett, H. H. and Myers, K. J. *Foundation of Image Science*. John Wiley & Sons (2004).
7. Vaishnav, J. Y., Jung, W. C., Popescu, L. M., Zeng, R. and Myers, K. J. *Objective assessment of image quality and dose reduction in CT iterative reconstruction*. Med. Phys. **41**, 071904 (2014).
8. Ott, J. G., Ba, A., Racine, D., Ryckx, N., Bochud, F. O., Alkadhi, H. and Verdun, F. R. *Patient exposure optimisation through task-based assessment of a new model-based iterative reconstruction technique*. Radiat. Prot. Dosim. DOI:10.1093/rpd/ncw019.
9. International Commission on Radiation Units and Measurements. *Receiver operating characteristic analysis in medical imaging*. ICRU Report 79. Journal of the ICRU 8(1) (2008).
10. International Electrotechnical Commission. International Standard IEC 60601-2-44. *Medical Electrical Equipment – Part 2-44: Particular Requirements for the Basic Safety and Essential Performance of X-Ray Equipment for Computed Tomography*. 3rd edn. IEC (2009).
11. Li, K., Garrett, J., Ge, Y. and Chen, G.-H. *Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance*. Med. Phys. **41**, 071911 (2014).

Towards a standardization of image quality in abdominal CT: Results from a multicentre study

Damien Racine[1], Nick Ryckx[1], Alexandre Ba[1], Fabio Becce[2], Anais Viry[1], Francis R. Verdun*[2] and Sabine Schmidt*[2]

[1] Institute of Radiation Physics – Lausanne University Hospital, Rue du Grand-Pré 1 – 1007 Lausanne – Switzerland

[2] Department of Diagnostic and Interventional Radiology – Lausanne University Hospital, Rue du Bugnon 46 – 1011 Lausanne – Switzerland

*F.R. Verdun and S. Schmidt contributed equally to this work

Abstract

Purpose: To highlight the spread of the diagnostic information available in CT images for 68 different CT units, although the selected respective protocols aim at answering the same clinical question.

Methods: An anthropomorphic abdominal phantom (QRM, Moehrendorf, Germany) with two optional rings was scanned on 68 CT machines using local clinical acquisition parameters of the portal phase for the detection of focal liver lesions. Low contrast detectability (LCD) was objectively assessed with a Channelized Hotelling Observer (CHO) using the receiver operating characteristic (ROC) paradigm. For each lesion size the area under the ROC curve (AUC) was calculated and considered as a figure of merit. The CTDIvol was used to indicate the dose exposure.

Results: The median $CTDI_{vol}$ used for acquisitions was 5.8 mGy, 10.5 mGy and 16.[illegible] mGy for the small, medium and large phantoms, respectively. The median AUC obtained from acquisitions was 0.96, 0.90 and 0.83 for the small, medium and large phantoms, respectively.

Conclusions: Our study shows that a standardization initiative could be launched and ensured comparable diagnostic information for a well defined clinical question. We thus propose the clinical image quality level as a starting point for the optimization process in diagnostic quality rather than the Diagnostic Reference Levels.

1. Introduction

In diagnostic radiology, computed tomography (CT) contributes to a major part of the public exposure which leads to a public concern due to potential cancer induction risks [1] [2] [3] [4]. In this context, many initiatives have been launched to avoid unnecessary or useless exposures, such as Image Gently or Image Wisely. The introduction of diagnostic reference levels (DRL) allowed, to a certain extent, to reduce the heterogeneity of the delivered dose exposure from one institution to another [5]. However, the DRL provided for CT examinations are generally defined as a function of an anatomical region which is certainly a limitation, since a given anatomical region may not need the same image quality depending on the clinical question (e.g. head trauma vs. sinusitis). In addition, technological developments, such as the Automatic Tube Current Modulation (ATCM), dynamic beam collimation to reduce over ranging were proposed to drastically reducing patient exposure [5]. Finally, in the last ten years iterative reconstruction techniques (IR) have become increasingly popular as a mechanism to reduce CT dose exposure at CT while ensuring a good level of image quality. In general, IR allow drastic image noise reductions while keeping a reasonable level of spatial resolution compared to traditional filtered back projection techniques [6] [7] [8] [9] [10].

With the large number of IR solutions now proposed by multiple CT vendors, it has become crucial to systematically evaluate the dose reduction potential and the subsequent image quality resulting for each technique. These investigations have been made on clinical images as well as on phantoms images [11] [12] [13]. It appears that, in spite of the production of subjectively better looking images IR does not allow a full recovery of the detection of low contrast structures when the applied dose reductions are too high [14] [15]. Thus, when dealing with dose reductions by means of IR, the low contrast detectability should be systematically investigated using task-based image quality assessment methodologies. Given the large number of IR solutions proposed by manufacturers such image quality assessments should be performed using phantoms and objective methods. For such evaluations, the Federal Drug Administration (FDA) recommends the use of mathematical model observers as surrogate to human observers. Their outcomes provide image quality indictors as the $CTDI_{vol}$, or patient exposure indicators as the dose length product (DLP).

The aim of this study was to investigate the variability of the patient exposure and the image quality provided by a large number of centres for the same clinical question: presence or absence of a focal lesion in the abdomen of phantoms of different morphologies.

2. Materials and methods

2.1 Description of the phantom

An anthropomorphic abdomen phantom (QRM 401, Moehrendorf, Germany) which simulates the attenuation produced by an adult (Figure 1) was used for this study. This phantom represents a thin adult with a body mass index (BMI) of 20 $kg.m^{-2}$ (or a patient's weight around 50 kg). To [illegible] the patient's morphology two additional rings, one of medium and another one of large size, were added to the phantom to reach a BMI of 26 $kg.m^{-2}$ (patient's weight around [illegible]0 kg) and 35 $kg.m^{-2}$ (patient's weight around 120 kilograms). A module containing spherical lesions of 8, 6 and 5 mm in diameter and a contrast of 20 HU relative to the background was inserted in the center of the phantom. Only the sphere measuring 5 mm in diameter was used in this study since preliminary measurements showed that this size was the most critical in terms of AUC. in terms of AUC.

Figure 1: CT images of the anthropomorphic abdomen phantom. From left to right : small, medium and large phantom.

2.2 CT units investigated

The three abdomen phantom sizes were scanned on 68 CT scanners installed in 62 institutions (*Table 1*). The four major manufacturers were represented. GE and Philips represents 68 % of the CT units involved in this study (see Figure 2).

Table 1: 70 CT scanners involved in this study

	CT unit	Number
GE	BrightSpeed S	1
	Discovery CT 750 HD	2
	LightSpeed VCT	7
	LightSpeed16	1
	Optima CT520	2
	Optima CT580	1
	Optima CT660	7
	Revolution	1
	Revolution EVO	1
Philips	Brilliance 40	1
	Brilliance 64	6
	iCT 256	4
	Ingenuity Core 128	4
	Ingenuity CT	4
	Ingenuity Flex	2
	IQon - Spectral CT	1
Siemens	Perspective	1
	Sensation 64	2
	SOMATOM Definition AS	2
	SOMATOM Force	1
Toshiba	Activion16	4
	Aquilion	6
	Aquilion PRIME	6

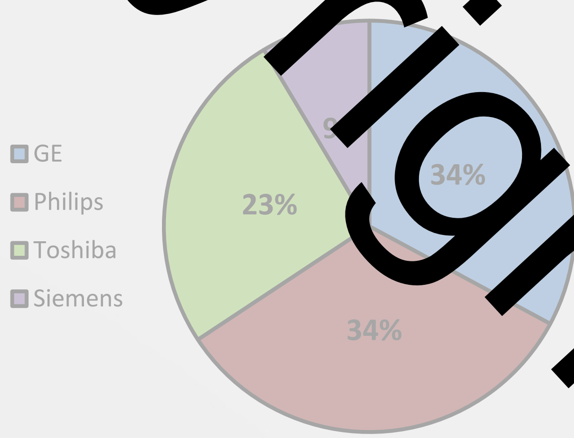

Figure 2: Distribution of the CT units among the four manufacturers

2.3 Acquisition protocols

The phantoms were always positioned at the isocenter of the CT units to maximize the performance of the ATCM [16] [17]. Image acquisitions were performed according to the local acquisition and reconstruction parameters of the portal phase used for focal liver lesion (FLL). To provide comparable spatial resolutions, the reconstructed field of views (FOV) were set to 320 mm, 370 mm and 420 mm for the small, medium and large size phantom, respectively. To ensure statistical robustness of the results the phantom was scanned ten times on each CT unit without changing the position between the different acquisitions [18].

2.4 Model observer - Channelized Hotelling Observer

A channelized Hotelling model observer (CHO) was used to assess the low contrast detectability. This model is called "channelized" because each image is passed through channels before the actual reading process, allowing for computation time sparing. These channels represent the spatial selectivity behaviour of human primary visual cortex (V1). CHO model is a linear and anthropomorphic model that computes a decision variable λ from the dot product between an image "g" and a template "w".

$$\lambda_{n,s,i} = w^T \bullet g_{n,s,i} \quad (1)$$

where T is the transpose operator; "n" represents the image category, signal-absent or signal-present, "s" represents the lesion size and "i" the image number. The template "w" takes into account the statistical knowledge of noise by computing the covariance matrix "K" from the images containing no signal seen through the channels. The template "w" also takes into account the signal by computing a theoretical signal that represents the different lesions. In this study a signal template, g_{Theo} is created using a simulated 2D Gaussian curve with a FWHM 5 mm, thus representing a theoretical signal of 5 mm in diameter.

$$\mathbf{w}_{\mathrm{CHO}} = \left(\mathbf{K}_{\mathrm{v/n}}\right)^{-1} \mathbf{g}_{\mathrm{Theo}} \quad (2)$$

We used ten dense of difference of Gaussian (D-DOG) channels because these channels are known to mimic human detection [19] [20] [21]. Each channel is given by the following formula:

$$C_j(\rho) = e^{-\frac{1}{2}\left(\frac{\rho}{Q\sigma_j}\right)^2} - e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2} \quad (3)$$

where ρ is the spatial frequency, σ_j the standard deviation of each channel and Q is the filter bandwidth. Each σ_j value is given by $\sigma_j = \sigma_0 \alpha_{j-1}$. The parameters generally used are: $\sigma_0 = 0.005$, $\alpha = 1.4$ and Q = 1.67 [19].

2.4.1 Performance measurement

From each image, the CHO model computes a decision variable, λ. After that, using both the decision variable distribution for noise only and signal images, a ROC study was performed to assess the image quality. The area under the curve (AUC) was used as figure of merit for image quality and the displayed $CTDI_{vol}$ was used as figure of merit for the dose exposure. The average and standard deviation of the AUC were estimated by performing a bootstrap method [22]. In practice, the model performed 500 ROC experiments for each category.

In addition, the correlation between the AUC of the different size was calculated using the Pearson coefficient.

3. Results

3.1 $CTDI_{vol}$ in terms of patient size

As expected, the $CTDI_{vol}$ increased with the size of the phantom. Due to the selection of the locally implemented protocol for each units, the $CTDI_{vol}$, varied significantly for a given phantom size. For the small phantom the range of $CTDI_{vol}$ was [illegible] mGy with a median $CTDI_{vol}$ equal to 5.8 mGy and a third quartile value equal to [illegible] mGy. For the medium phantom the range of $CTDI_{vol}$ was [illegible].5-34.2 mGy with a median $CTDI_{vol}$ of [illegible] mGy and a third quartile value of 13.4 mGy, and for the largest phantom size the range of $CTDI_{vol}$ was 8.6-34.2 mGy with a median $CTDI_{vol}$ of [illegible] mGy and a third quartile value of 20.9 mGy (see Figure 3).

Figure 3: CTDI$_{vol}$ obtained for the three phantoms sizes as a function of the clinical protocol settings. The median is represented by the black line in the middle of the colour rectangles. The bottom edge of the rectangle corresponds to the 1st quartile and the top edge to the 3rd quartile. The bottom line represents the 5th percentile and the top line the 95th percentile. The red dots outside these two lines represent the outliers.

3.2 Image quality for abdomen protocols

3.2.1 Small phantom

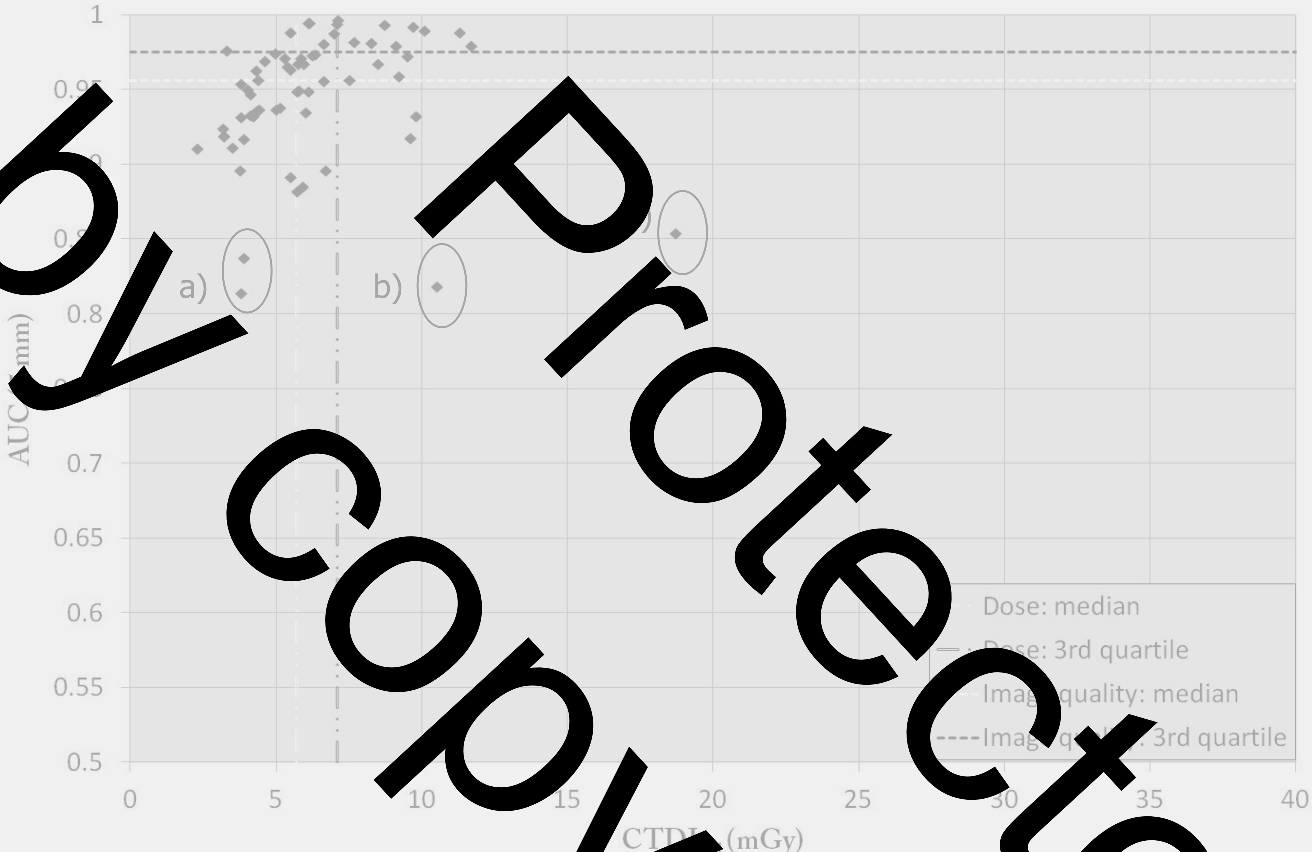

Figure 4: AUC as a function of CTDI$_{vol}$ for the small phantom

For the small phantom size, in spite of the use of relatively low CTDI$_{vol}$ values (median CTDI$_{vol}$ equals to 5.80 mGy) in comparison to the DRL values (15 mGy), an excellent image quality level was obtained for most of the centers: the median value of the AUC was equal to 0.96 and the third quartile was equal to 0.97 with several centers reaching an AUC superior to 0.99. Only three centers had an AUC lower than 0.85. Two protocols (a) used a too low dose level for the CT considered to obtain a good image quality. The slice thickness used for centers b and c was equal to 5mm and can't be compatible to correctly detect a lesion of 5 mm in diameter due to partial volume effect even if the dose level was high.

3.2.2 Medium phantom

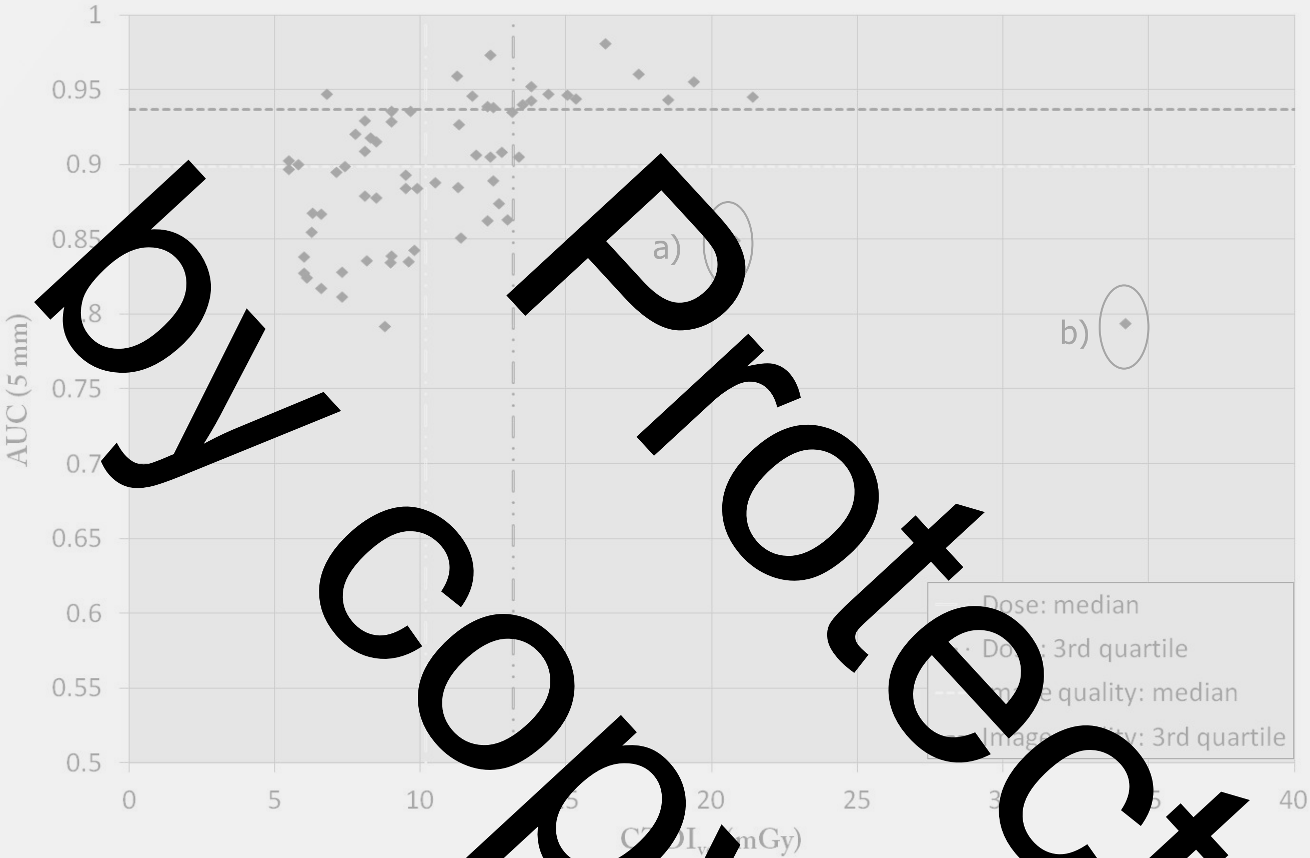

Figure 5: AUC as a function of $CTDI_{vol}$ for the medium size phantom

For the medium phantom size the median $CTDI_{vol}$ was equal to 10.2[illegible] mGy with a median value of the AUC reaching 0.90 (the third quartile value was equal to 0.94). With this phantom size almost a quarter of the centers had an AUC lower than 0.850. As already seen in Fig.4, the dose levels were relatively to significantly high in three protocols (Fig.5, a and b) caused by a suboptimal reconstruction slice thickness (5 mm). These images had been acquired on the oldest CT machines included in this study that had been introduced on the market during the years 2007 – 2009.

3.2.3 Large phantom

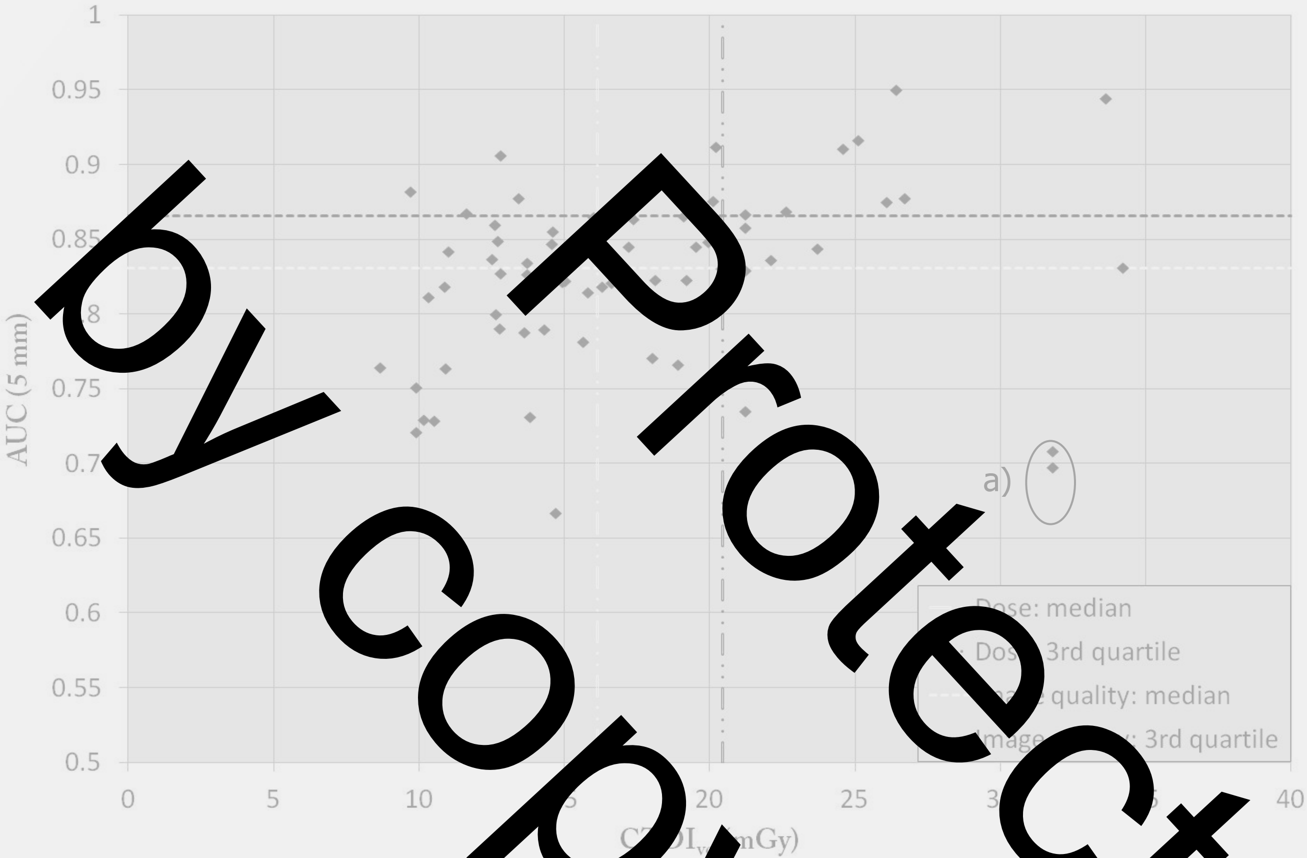

Figure 6: AUC as a function of $CTDI_{vol}$ for the large phantom.

For the large phantom size the median CTDIvol was equal to 16.14 mGy, with a median value of the AUC reaching 0.83 (the third quartile value was equal to 0.87). With this phantom size almost a quarter of the centers had an AUC lower than 0.80. Like for the other phantom sizes, the image quality was significantly low in two protocols (a) due to partial volume effects, again associated with the use of one of the oldest CT units.

3.2.4 Image quality correlation

	Size S	*Size M*	*Size L*
Size S	1		
Size M	0.39	1	
Size L	0.33	0.49	1

Table 2: Correlation matrix based on AUC obtained with ATCM that maintained absolute noise levels close to the target values.

	Size S	Size M	Size L
Size S	1		
Size M	0.49	1	
Size L	0.40	0.58	1

Table 3: Correlation matrix based on AUC obtained with ATCM that maintained a constant level of overall diagnostic quality for all patient sizes relating to a reference image

In our study we included two types of ATCM: one type of ATCM for which the user has to choose a noise level and another type of ATCM for which the user only introduces a reference image load (in mAs). For the first type of ATCM the correlation of the AUC values between the different sizes varied from 0.33 between the size S and L to 0.49 between the size M and L (Table 2). For the second type of ATCM the correlation of the AUC values varied from 0.40 between the size S and L to 0.58 between the size M and L (Table 3). In summary, it appears that there was a poor correlation between the level of image quality and the size of the phantoms.

4. Discussion

In this study, the image quality was assessed by applying a task-based criterion on three abdominal phantom sizes and assessing low-contrast detectability by means of a CHO model observer. In the figures 4 - 6 four quadrants created from the median dose and median AUC were proposed to discriminate the various centers. The top left quadrant contains the CT units which machines and clinical protocols provide a high level of image quality in a low dose range. For the chosen task-based image quality criterion these centers produced the best performances. In the bottom left quadrant, the protocols led to lower image quality levels in spite of comparable dose exposures. This was sometimes due to the use of sub-optimal ATCM settings or a sub-optimal reconstruction slice thickness (Figure 5). CT units placed in the bottom right quadrant are these for which an optimization process should be initiated since low image quality levels are obtained in spite of high dose levels.

It is worth mentioning that with the increase in phantom size the variability of results in terms of the image quality and dose increased. As expected, with the increase of the phantom size the top right quadrant required a higher dose level to keep an excellent diagnostic accuracy (using the AUC parameter as surrogate).

The overall results show that for a given protocol a large variability of image quality is provided and that this spread is even more important with increasing patient weight (or BMI) although the clinical question to be answered remains the same. This is certainly due to the usually adopted strategy of optimization with based on national DRL. Indeed, for many years the use of the DRL concept has allowed to reduce the variability of dose exposure in the practice using reference levels that were defined for specific anatomical regions. The application of the DRL concept has been widely used, and it allowed a national homogenisation of the patient's exposure. This optimization strategy should be now questioned. With the use of iterative reconstructions reasonably "good looking" CT images can be obtained in a larger dose range than when using the standard FBP reconstruction process. However, the amount of information contained in images remains related to the number of photons used to produce these images. When one wants to justify a given CT examination, one should adapt the imaging protocol in that way that one is able answer one or several clinical questions, that correspond to the indication of the examination. The challenge is now to establish a direct link between the different clinical tasks and image quality metrics. These task-based image quality criteria (i.e. contrast detectability) should also be defined as a function of the patients' morphology. The outcome of this study demonstrates that a standardization process concerning the image quality requirements as a function of a given clinical indication should be initiated for different patient sizes. It appears that the image quality is relatively homogeneous with the small phantom, but when the phantom size was increased the image quality tended to be different between the different institutions. This would not only ensure that all patients benefit from the same diagnostic potential inherent in their CT examinations but could also limit the "blind" dose reductions which may impair the diagnostic image quality.

Although those results assess image quality on phantoms measurements using clinical protocols, our study presented some limitations. First of all, we did not consider the management of the contrast medium injection. However, the latter is an important part of the optimization protocol and has a undeniable impact on the contrast enhancement of the organs and vessels, and thus on the diagnostic result. The other limitation was due to the use of a homogeneous phantom that simplifies the objective assessment of image quality but remains far from the actual clinical situation but it is a good starting point for an optimization process. Finally, the largest phantom was made only of muscular tissue which is not fully representative of obese patients. This increased the performance requirements for the CT units. The largest rings added to the core of the phantom should have contained muscular tissue and fat and not only muscular tissue. Another limitation concerned the material which composed the lesion present in the phantom. When we investigated the portal venous phase on an abdominal protocol - which is de facto an injected phase - it would have been more clinically relevant to use a phantom with targets containing a high Z material such as iodine.

A method for benchmarking CT abdominal protocols has been applied with three different phantom sizes by evaluating 68 different CT units. This study demonstrates that the cooperation between the radiologists, radiographers, and the medical physicist, must be promoted to ensure that dose reductions do not lead to sub-optimal images that might impair the diagnostic quality. The aim of the previously introduced DRL concept was to reduce the variability of the patients' dose exposure but, as we could show in our study, they have revealed insufficient to ensure comparable image quality levels on different CT machines. We should now define a set of task-based image quality criteria related to well-defined clinical indications and work towards the standardization of image quality requirements. To establish these image quality requirements it is important to define in collaboration with radiologists the critical targets to be detected and at which AUC level we should use our standardized phantoms.

5. References

[1] A.-F. Perez, C. Devic, C. Colin, and N. Foray, "[The low doses of radiation: Towards a new reading of the risk assessment]," *Bull. Cancer (Paris)*, vol. 102, no. 6, pp. 527–538, Jun. 2015.

[2] J. D. Mathews *et al.*, "Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians," *BMJ*, vol. 346, p. f2360, May 2013.

[3] H. Hricak *et al.*, "Managing radiation use in medical imaging: a multifaceted challenge," *Radiology*, vol. 258, no. 3, pp. 889–905, Mar. 2011.

[4] D. J. Brenner, I. Shuryak, and A. J. Einstein, "Impact of reduced patient life expectancy on potential cancer risks from radiologic imaging," *Radiology*, vol. 261, no. 1, pp. 193–198, Oct. 2011.

[5] M. Rosenstein, "Diagnostic reference levels for medical exposure of patients: ICRP guidance and related ICRU quantities," *Health Phys.*, vol. 95, no. 5, pp. 528–534, Nov. 2008.

[6] A. K. Hara, R. G. Paden, A. C. Silva, J. L. Kujak, H. J. Lawder, and W. Pavlicek, "Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study," *AJR Am. J. Roentgenol.*, vol. 193, no. 3, pp. 764–771, Sep. 2009.

[7] A. C. Silva, H. J. Lawder, A. Hara, J. Kujak, and W. Pavlicek, "Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm," *AJR Am. J. Roentgenol.*, vol. 194, no. 1, pp. 191–199, Jan. 2010.

[8] Y. Sagara, A. K. Hara, W. Pavlicek, A. C. Silva, R. G. Paden, and Q. Wu, "Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients," *AJR Am. J. Roentgenol.*, vol. 195, no. 3, pp. 713–719, Sep. 2010.

[9] F. Pontana *et al.*, "Chest computed tomography using iterative reconstruction vs filtered back projection (Part 2): image quality of low-dose CT examinations in 80 patients," *Eur. Radiol.*, vol. 21, no. 3, pp. 636–643, Mar. 2011.

[10] Y. Funama *et al.*, "Combination of a low-tube-voltage technique with hybrid iterative reconstruction (iDose) algorithm at coronary computed tomographic angiography," *J. Comput. Assist. Tomogr.*, vol. 35, no. 4, pp. 480–485, Aug. 2011.

[11] S. T. Schindera *et al.*, "Iterative Reconstruction Algorithm for CT: Can Radiation Dose Be Decreased While Low-Contrast Detectability Is Preserved?," *Radiology*, vol. 269, no. 2, pp. 511–518, Nov. 2013.

[12] K. T. Flicek, A. K. Hara, A. C. Silva, Q. Wu, [illegible], and C. D. Johnson, "Reducing the radiation dose for CT colonography using adaptive statistical iterative reconstruction: A pilot study," *AJR Am. J. Roentgenol.*, vol. 195, no. 1, pp. 126–131, Jul. 2010.

[13] D. Marin *et al.*, "Low-tube-voltage, high-tube-current multidetector abdominal CT: improved image quality and decreased radiation dose with adaptive statistical iterative reconstruction algorithm--initial clinical experience," *Radiology*, vol. 254, no. 1, pp. 145–153, Jan. 2010.

[14] K. L. Dobeli, S. J. Lewis, S. R. Meikle, D. L. Thiele, and P. C. Brennan, "Noise-reducing algorithms do not necessarily provide superior dose optimisation for hepatic lesion detection with multidetector CT," *Br. J. Radiol.*, vol. 86, no. 10[illegible], p. 20120500, Mar. 2013.

[15] M. E. Baker *et al.*, "Contrast-to-noise ratio and low-contrast object resolution on full- and low-dose MDCT: SAFIRE versus filtered back projection in a low-contrast object phantom and in the liver," *AJR Am. J. Roentgenol.*, vol. 199, no. 1, pp. 8–18, Jul. 2012.

[16] M. A. Habibzadeh, M. R. Ay, A. R. K. Asl, H. Ghadiri, and H. Zaidi, "Impact of miscentering on patient dose and image noise in x-ray CT imaging: phantom and clinical studies," *Phys. Medica PM Int. J. Devoted Appl. Phys. Med. Biol. Off. J. Ital. Assoc. Biomed. Phys. AIFB*, vol. 28, no. 3, pp. 191–199, Jul. 2012.

[17] S. T. Schindera *et al.*, "Effect of automatic tube current modulation on radiation dose and image quality for low tube voltage multidetector row CT angiography: phantom study," *Acad. Radiol.*, vol. 16, no. 8, pp. 997–1002, Aug. 2009.

[18] D. Racine, A. H. Ba, J. G. Ott, F. O. Bochud, and F. R. Verdun, "Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer," *Phys. Medica Eur. J. Med. Phys.*, vol. 32, no. 1, pp. 76–83, Jan. 2016.

[19] C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A*, vol. 18, no. 3, pp. 473–488, Mar. 2001.

[20] H.-W. Tseng, J. Fan, M. A. Kupinski, P. Sainath, and J. Hsieh, "Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers," *Med. Phys.*, vol. 41, no. 7, p. 071910, Jul. 2014.

[21] S. Leng, L. Yu, Y. Zhang, R. Carter, A. Y. Toledano, and C. H. McCollough, "Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain," *Med. Phys.*, vol. 40, no. 8, p. 081908, Aug. 2013.

[22] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. CRC Press, 1994.

Task-based assessment of impact of multiplanar reformations on objective image quality in iterative reconstruction in computed tomography

Damien Racine[1], Julien G. Ott[1], Pascal Monnin[1], David C. Rotzinger[2], Patrick Omoumi[2], Francis R. Verdun[1], Fabio Becce[2]

[1] Institute of Radiation Physics, CHUV, Lausanne, Switzerland
[2] Department of Diagnostic and Interventional Radiology, CHUV, Lausanne, Switzerland

Introduction

Computed tomography (CT) images are traditionally acquired, reconstructed and analyzed in the axial plane. However, in several clinical situations, CT images need to be analyzed in the coronal and/or sagittal planes, particularly in cardiovascular, thoracic and musculoskeletal imaging [Fang et al 2015]. With the arrival of CT 3D imaging it has become possible today to acquire and reconstruct CT images in all reconstruction planes [Dalrymple et al 2007]. Several authors [Long 2010 *et al;* Rydberg *et al* 2007; Von Falck *et al* 2011] already investigated the impact of sagittal and coronal reformatting for various clinical CT acquisition conditions. Most of the work conducted so far was however only done for classical filtered-back projection (FBP) algorithms and using subjective image quality assessment methods like visual grading analysis made by human observers.

The recent implementation of iterative reconstruction (IR) techniques in clinical CT protocols has helped to significantly reduce radiation dose, but with a potential change in image quality. The impact of the various manufacturer-specific IR techniques on image quality has already been extensively studied in the axial plane using both visual grading analysis and physics methods fully adapted to IR [Schindera *et al* 2011, Mieville *et al* 2012; Ott *et al* 2014]. However, IR impact on the coronal and sagittal reconstruction planes has not yet been done using objective and fully IR-adapted methods.

Our work will focus on comparing image quality in all three reconstruction planes using objective assessment methods adapted to IR. We focus on a great variety of algorithms designed by a single manufacturer.

Material and Methods

Phantom

In order to evaluate key CT image quality parameters like image noise or spatial resolution, we built our own phantom, containing a 10-cm-diameter cylinder made of Teflon® (polytetrafluoroethylene), low-density Polyethylene and Plexiglas® (polymethyl methacrylate) and centered in the middle of air (figure 1). Among those materials, Teflon® revealed itself to be particularly useful since its contrast with background was close to the contrast between cortical bone and air, or articular cartilage and diluted iodine (2000 HU). This experimental paradigm enabled us to produce phantom images similar to clinical CT images in various musculoskeletal CT settings, such as the search of fracture lines in cortical bone after trauma, or articular cartilage defects and labral/fibrocartilage tears in CT arthrography.

Figure 1: a) Custom-made image quality phantom containing a 10 cm-diameter cylinder made of Teflon®, low-density Polyethylene and Plexiglas® (from left to right). **b)** A slice of our custom made phantom.

Protocol

Our custom-made image quality phantom was scanned on a GE HD 750 CT scanner (GE Healthcare, USA). The acquisitions were performed with a protocol similar to the one used in clinical routine for patients undergoing CT arthrography, that is today with a $CTDI_{vol}$ of 7.5 mGy, a tube voltage of 120 kVp, a pitch equal to 0.5 and a field of view of 200 mm with a matrix size of 512 × 512 pixels. The acquired data sets were then reconstructed in the axial, sagittal and coronal planes, using a nominal slice thickness of 0.625 mm, and four different manufacturer-specific reconstruction algorithms: the classical filtered-back projection (FBP), the adaptive statistical iterative reconstruction (ASIR) at a percentage of 50 %, and the Adaptative statistical iterative reconstruction –V (ASIR-V) at a percentage of 50 %,, with this three algorithms the GE bone kernel was used. Moreover, the GE model-based iterative reconstruction; “VEO” algorithm was also used. Images were reconstructed with VEO 2.0 that was only compatible with the standard kernel. However, new presets are provided in the VEO 3.0 version, including resolution preference (RP**). The preset index exactly describes the expected resolution improvement as measured on a GE performance phantom. RP05 implies 5% higher resolution compared to standard preset. Also, the image model and noise model of the algorithm were improved to make the noise covariance more isotropic in all three dimensions. With VEO 3.0 images were reconstructed with the preset RP 05 and RP 20.

Physics Metrics

Iterative reconstruction (IR) algorithms are known to be highly non-linear and therefore to introduce a dependency of the image contrast and noise over the spatial resolution of the image [Thibault *et al* 2007; Richard *et al* 2012]. In order to overcome those problems, spatial resolution was investigated through an object-specific modulation transfer function (MTF) which we referred to as target transfer function (TTF) [Richard *et al* 2012]. In our case, this object-specific metric was obtained using the contrast variation between the circular edge of the Teflon® cylinder and the contrast product around it. Main mathematical steps, as well as extensive details and explanations on the methodology can be found in the paper of Ott et al. [Ott *et al*, 2014].

The image noise was investigated within a 6-cm long Plexiglas region of our phantom. NPS were calculated based on 70 image slices of the homogenous contrast-media region containing regions of interests (ROIs) of [illegible] pixels. A radial-averaged NPS was obtained based on the guidelines described in the ICRU report 54 [ICRU 54]. Extensive details on the NPS metric as well as elements to perform its computation can be found in Miéville's work [Miéville *et al* 2011].

We integrated those two metrics into a model observer in order to yield an objective assessment of the image quality in the three reconstruction planes when using different reconstruction algorithms. Model observers rely on the concept of task-based assessment in order to assess how well the desired information can be extracted from the image [Barrett and Meyers 2004]. In our case, we used an updated non-prewhitening with eye filter (NPWE) model observer to compute the d' index of simulated articular cartilage defects of variable width (0.5, 1, 1.5, and 2 mm, respectively) using Equation 1 [Brunner and Kyprianou 2011; Ott *et al* 2014].

$$d' = \frac{\sqrt{2\pi}\,\Delta HU \int_0^{f_{Ny}} S^2(f)TTF^2(f)\,VTF^2(f)f\,df}{\sqrt{\int_0^{f_{Ny}} S^2(f)TTF^2(f)\,NPS(f)\,VTF^4(f)f\,df}} \qquad (1)$$

f_{Ny} is the Nyquist frequency, ΔHU is the contrast difference, $VTF(f)$ is the visual transfer function of the human eye [Burgess 1994] and $S(f)$ is the Fourier transform of the input signal (in our case, $S(f) = \frac{R}{f} J_1(2\pi Rf)$, J_1 being a Bessel function of the first kind).

Results

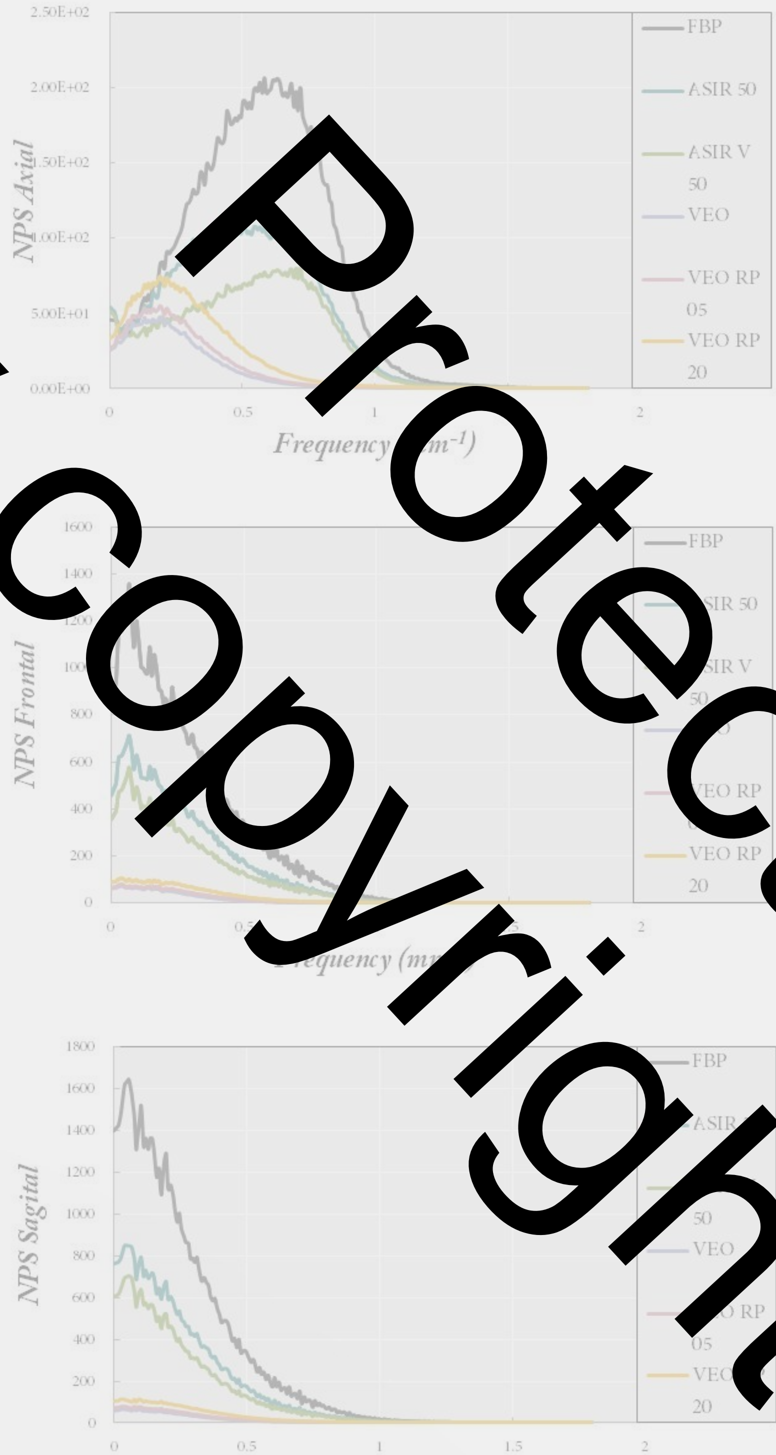

Figure 2: NPS curves computed for FBP, ASIR, ASIR-V, VEO 2.0 and VEO 3.0 reconstruction algorithms in the three reconstruction planes.

Figure 3: TTF curves computed for FBP, ASIR, ASIR-V, VEO 2.0 and VEO 3.0 reconstruction algorithms in the three reconstruction planes.

FBP, ASIR 40 & 80, VEO 2.0

Variations in NPS as a function of the image reconstruction algorithm and reconstruction plane were observed (Figure 2). Those changes in the shape of the NPS curves indicate a modification of the overall image texture. For the FBP, ASIR and ASIR-V algorithms coronal and sagittal reformatted images display a pick frequency at lower frequencies than axial images. NPS of images reconstructed in the axial plane, we observed that the noise level slightly increases when switching from the standard kernel to the RP05 kernel to the RP20 kernel for VEO algorithm.

In the three reconstructed plane, FBP algorithm had the higher noise level. Switching from FBP to iterative algorithm did however yield to a consequent noise decrease, especially in the low frequencies range. In coronal and sagittal-reformatted images, we noted an increase in image noise compared with the axial plane. This trend was even slightly more pronounced in the sagittal plane.

TTF curves (figure 3) in the axial plane indicated that an improvement of the spatial resolution was perceived when using ASIR or ASIR-V instead of FBP. On the contrary, the use of VEO produced a decrease of the resolution for the whole frequency range of the image. The behavior observed in the coronal and sagittal planes was completely different because the higher TTF curve is obtained with VEO algorithm. Moreover the plans RP05 and RP20 did not increase the spatial resolution whatever the reconstruction plane. For all algorithms, the TTFs curves suggest that, a decrease of the spatial resolution happens when switching from axial to coronal or sagittal reconstruction plane.

Our results suggest that full model based algorithm like VEO improves the detectability in comparison to the other algorithms. Moreover, major changes in the detectability as estimated by the NPWE model observer in the sagittal and the coronal plane compared with the axial plane when images are reconstructed with FBP or statistical iterative algorithms (Table 1). However, we observed a constant detectability in all reconstruction planes when using VEO instead of FBP, ASIR or ASIR-V demonstrating that the use of this MBIR algorithm could help to improve the diagnosis accuracy. The detectability using a standard kernel with VEO yields a higher detectability than the RP05, which itself produces higher detectability than the RP20 kernel. This trend is due to a higher noise level when using sharper resolution kernels and which is not fully compensated by the signal increase these kernels also generate.

FBP				
Size	0.5	1	1.5	2
Axial	1.51	5.03	8.99	12.91
Front	0.40	1.38	2.52	3.56
Sagittal	0.36	1.26	2.31	3.25

ASIR 50				
Size	0.5	1	1.5	2
Axial	[illegible]	6.88	11.75	16.11
Front	0.59	[illegible]	3.63	5.03
Sagittal	0.55	[illegible]	3.35	4.63

ASIR-V 50				
Size	0.5	1	1.5	2
Axial	2.59	[illegible]	[illegible]	20.81
Front	0.68	2.35	4.24	5.91
Sagittal	0.63	2.16	[illegible]	5.37

VEO				
Size	0.5	1	1.5	2
Axial	4.70	14.09	21.59	26.52
Front	[illegible]	12.16	19.83	[illegible]
Sagittal	3.53	11.40	19.07	[illegible]

VEO RP 05				
Size	0.5	1	1.5	2
Axial	4.03	12.25	[illegible]	23.70
Front	3.12	9.87	16.10	20.55
Sagittal	2.87	9.26	15.42	[illegible]

VEO RP 20				
Size	0.5	1	1.5	2
Axial	3.06	9.38	14.82	[illegible]
Front	2.35	7.51	12.38	[illegible]
Sagittal	2.16	7.03	11.83	[illegible]

Table 1: d’ computed for FBP, ASIR, ASIR-V, VEO 2.0 and VEO 3.0 reconstruction algorithms in the three reconstruction planes for simulated cartilaginous lesion of 0.5 mm, 1.0 mm, 1.5 mm and 2.0 mm.

Discussion

Our results indicate that in our paradigm, using ASIR algorithms instead of FBP does not bring any major improvement regarding the SNR. This result was already observed in other studies focusing on the axial plane exclusively [Racine et al 2015] and remained true in the three different reconstruction planes. Yet, the use of MBIR like VEO 2.0 happened to lead to more conclusive effects on the SNR values. In the axial plane, a consequent noise reduction was indeed observed with VEO 2.0 compared to FBP, whereas an increase of the noise was observed in the coronal and sagittal planes when switching from FBP to VEO 2.0. In the meantime, resolution found out to be improved with VEO 2.0 when switching from the axial to the other planes. Eventually, using VEO 2.0 instead of FBP led to a significant SNR increase in all reconstruction planes. The previously stated results regarding noise and resolution behavior remained true and even reinforced when switching from FBP to VEO 3.0. Results suggest that the 3.0 version leads to a higher SNR in all three reconstruction planes than the 2.0. In the end, switching reconstruction planes and algorithms has repercussions on image quality, so that lesion detection and characterization by the radiologist are therefore modified.

Image quality depending on the reconstruction plane was already investigated by several authors [Rydberg *et al* 2007; [illegible]gla Lo[illegible] *et al* 2010] who successfully demonstrated that multi-planar reconstructions could improve diagnostic accuracy and interpretation speed. These studies were however conducted relying on subjective image quality assessment methods, i.e. scoring of the images by human observers in this case. If this kind of methods has already demonstrated its great potential to assess the quality of CT images, there is also a need for objective methods to perform this task. Thus, other authors chose to adopt those objective methods in order to study the impact of the reconstruction plane over the image quality [Von Falck *et al* 2011] and to conclude on the usefulness of the multi-planar reconstructions. Most of the objective methods used to characterize image quality however often consist in the computation of pixels' standard deviation, or of contrast-to-noise ratio. The study therefore provided useful and novel results regarding the quality of images reconstructed in different planes, but the outcome of these simple metrics is systematically improved when working with IR, because those algorithms integrate the knowledge of the noise statistics to penalize pixel variations and reduce noise.

This therefore enlightens the need to develop more elaborated tools, which are in compliance with the requirements of IR techniques, and we believe that methods such as task-based assessment could represent an efficient way out of the problem. Some authors like Guggenberger [Guggenberger *et al* 2013] already tackled the problematic of the multi-planar reconstruction using this kind of method. The study in question was however limited to FBP algorithms and we believe that IR algorithms, now that they are widely used in clinical routine should also be assessed this way. Our study was therefore conducted in order to objectively assess several types of IR, including some recently released MBIR algorithm, in different planes and using a task-based assessment method.

There are however some limitations to our work that have to be mentioned, the first one being that no image quality assessment was performed in our experimental paradigm. Indeed, if it is true that this kind of assessment is subject to some drawback like the subjectivity of the observer, they still represent the gold standard and comparing our objective results to human observers could be interesting. The second limitation comes from the fact that image quality assessment was performed but the diagnostic accuracy was not evaluated. To date, most of the published literature on IR techniques has reported reduction in radiation dose while maintaining diagnostic image quality, but very few have evaluated the impact of IR techniques on the diagnostic performance.

The results we obtained show that MBIR are indubitably the reconstruction algorithms leading to the highest SNR level in every reconstruction planes. The [illegible].0 version even surpasses the 3.0 version thanks to its possibility to change the reconstruction kernel and its results leading to higher SNR values.

Acknowledgment

/

References

[Barrett and Meyers 2004] H. H. Barrett and K. J. Myers, “*Foundations of Image Science.*” Wiley, New York, 2004.

[Brunner and Kyprianou 2011] C. C. Brunner and I. S. Kyprianou, *"Material-specific transfer function model and SNR in CT"* 2013 PhysMed Biol **58** 7447-61.

[Burgess 1994] A. E.Burgess, *"Statistically defined backgrounds: performance of a modified non prewhitening observer model"* 1994 JOSA A **11(4)**:1237-42.

[Dalrymple *et al* 2007] N. C. Dalrymple, S. R. Prasad, F. M. El-Merhi, K. N. Chintapalli, “Price of isotropy in multidetector CT.” 2007 Radiographics. **27**(1):49-62.

[Fang *et al* 2015] Q. Fang , F. Chen, A. Jiang, Y. Huang and K. Deng, *“Computed tomographic angiography of the superficial cerebral venous anastomosis based on volume rendering, multi-planar reconstruction, and integral imaging display.”* 2015 Australas Phys Eng Sci Med.

[Guggenberger *et al* 2013] R. Guggenberger, S. Winklhofer, J. V. Spiczak, G. Andreisek and H. Alkadhi, *“ In vitro high-resolution flat-panel computed tomographic arthrography for artificial cartilage defect detection: comparison with multidetector computed tomography.”* 2013 Invest Radiol. **48(8)**:614-21.

[ICRU 54] International Commission on Radiation Units and Measurements, “*Medical Imaging: The Assessment of Image Quality*.” 1996 Report 54 ICRU Publications, Bethesda, MD.

[Judy 1976] P. F. Judy, *"The line spread function and modulation transfer function of a computed* *tomographic scanner."* 1976 Med Phys **3**:233–6.

[Miéville *et al* 2011] F. A. Miéville, F. Gudinchet, E. Rizzo, P. Ou, F. Brunelle, F. Bochud, and F R. Verdun, *"Pediatric cardiac CT examinations: impact of the iterative reconstruction method (ASIR) on image quality – a preliminary study"* 2011 Pediatr Radiol **41**, 1154–64.

[Mieville *et al* 2012] F. A. Miéville, F. Gudinchet, F. Brunelle, F. O. Bochud and F. R. Verdun, *"Iterative reconstruction methods in two different MDCT scanners: physical metrics and 4-alternative forced-choice detectability experiments--a phantom approach."* 2012 Phys Med. **29(1)**:99-110.

[Ott *et al*, 2014] [Ott, 2014] J. G. Ott, F. Becce, P. Monnin, S. Schmidt, F.O. Bochud and F. R. Verdun, *"Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms."* 2014 Phys Med Biol **59(15)** 4047-[illegible].

[Racine *et al* 2015] D. Racine, A. H. Ba, J. G. Ott, F. O. Bochud and F. R. Verdun, "Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer." 2015 Phys Med. S1120-1797[illegible]5)0099[illegible]7

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

[Rydberg *et al* 2007] J Rydberg, K. Sandrasegaran, R. D. Tarver, M. S. Frank, E. J. Conces, R. H. Choplin, *"Routine Isotropic Computed Tomography Scanning of Chest Value of Coronal and Sagittal Reformations."* 2007 Invest Radiol 42(1):23-[illegible].

[Schindera *et al* 2011] S. T. Schindera, L. Diedrichsen, H. C. Müller, O. Rusch, D. Marin, B. Schmidt, R. Raupach, P. Vock and Z. Szucs-Farkas, *"Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom study."* 2011 Radiology **260(2)**:454-62.

[Singla Long *et al* 2010] S. S. Long, P. T. Johnson, Karen M. Horton, and Elliot K. Fishman, *"Are Multiplanar Reconstructions Necessary in Routine Body Computed Tomography Practice?: What Is the Published Evidence?"* 2010 J Comput Assist Tomogr **34**: 689-[illegible]8.

[Thibault *et al* 2007] [Thibault, 2007] J. B. Thibault, K. D. Sauer, C. A. Bouman and J. Hsieh, *"A three-dimensional statistical approach to improved image quality for multislice helical CT"* 2007 Med Phys **34**, 4526–4544.

[Von Falck *et al* 2011] C. von Falck, P. Hollmann, T. Rodt, S. Waldeck, B. Meyer, F. Wacker and H. O. Shin, *"Influence of multiplanar reformations on low-contrast performance in thin-collimated multidetector computed tomography."* 2011 Invest Radiol. **46(10)**:632-8

Physica Medica 31 (2015) 823–843

Contents lists available at ScienceDirect

Physica Medica

journal homepage: http://www.physicamedica.com

Review Paper

Image quality in CT: From physical measurements to model observers

F.R. Verdun [a,*,1], D. Racine [a,1], J.G. Ott [a], M.J. Tapiovaara [b], P. Toroi [b], F.O. Bochud [a], W.J.H. Veldkamp [c,d], A. Schegerer [e], R.W. Bouwman [d], I. Hernandez Giron [c], N.W. Marshall [f], S. Edyvean [g]

[a] Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland
[b] STUK-Radiation and Nuclear Safety Authority, PO Box 14, FIN-00881 Helsinki, Finland
[c] Department of Radiology, Leiden University Medical Center, C2-S, PO Box 9600, 2300RC Leiden, The Netherlands
[d] Dutch reference Centre for Screening (LRCB), Radboud University Medical Centre, PO Box 6873, 6503 GJ Nijmegen, The Netherlands
[e] Department of Radiation Protection and Health External and Internal Dosimetry, Biokinetics, Ingolstädter Landstr, 1, 85764 Neuherberg, Germany
[f] Katholieke Universiteit Leuven, Oude Markt 13, 3000 Leuven, Belgium
[g] Medical Dosimetry Group, Centre for Radiation Chemicals and Environmental Hazards, Public Health England, Didcot, UK

ARTICLE INFO

Article history:
Received 8 May 2015
Received in revised form 4 August 2015
Accepted 23 August 2015
Available online 12 October 2015

Keywords:
Computed tomography
Image quality
Patient dose optimisation
Model observer

ABSTRACT

Evaluation of image quality (IQ) in Computed Tomography (CT) is important to ensure that diagnostic questions are correctly answered, whilst keeping radiation dose to the patient as low as is reasonably possible. The assessment of individual aspects of IQ is already a key component of routine quality control of medical x-ray devices. These values together with standard dose indicators can be used to give rise to 'figures of merit' (FOM) to characterise the dose efficiency of the CT scanners operating in certain modes. The demand for clinically relevant IQ characterisation has naturally increased with the development of CT technology (detectors efficiency, image reconstruction and processing), resulting in the adaptation and evolution of assessment methods. The purpose of this review is to present the spectrum of various methods that have been used to characterise image quality in CT: from objective measurements of physical parameters to clinically task-based approaches (i.e. model observer (MO) approach) including pure human observer approach. When combined together with a dose indicator, a generalised dose efficiency index can be explored in a framework of system and patient dose optimisation. We will focus on the IQ methodologies that are required for dealing with standard reconstruction, but also for iterative reconstruction algorithms. With this concept the previously used FOM will be presented with a proposal to update them in order to make them relevant and up to date with technological progress. The MO that objectively assesses IQ for clinically relevant tasks represents the most promising method in terms of radiologist sensitivity performance and therefore of most relevance in the clinical environment.

Introduction

Diagnostic x-rays contribute to nearly 50% of the total annual collective effective dose of radiations from man-made and natural sources to the general population in western countries; computed tomography (CT) is the largest single source of this medical exposure.

The contribution of CT to collective dose has significantly increased in recent years and a considerable effort is required to control this trend and ensure that the benefits from the use of this technology outweigh the risks [1]. For example, in 2007–2008 the average dose per inhabitant, due to CT, was about 0.8 mSv in France and Switzerland, and about 0.7 mSv in Germany (as part of an average for all x-ray imaging of about 1.2 mSv and 1.7 mSv, respectively) [2–4]. An update of the French and German data showed that in 2012 the contribution of CT exposure has increased to approximately 1.15 mSv, with a similar increase shown in the last Swiss survey performed for 2013 [5].

In this context the radiation protection requirements in diagnostic radiology (justification of the examination and optimisation of the imaging protocol) need to be re-enforced. Justifying a CT scan is a clinical consideration and therefore will not be addressed in this work. However, the optimisation of a CT examination is achieved when image quality enables the clinical question to be answered whilst keeping patient radiation dose as low as reasonably possible. For this purpose the clinical question needs to be formulated as concretely as possible to enable a clear description of the image quality level required. To achieve this, appropriate and clinically

* Corresponding author. Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland. Tel.: +41 21 314 82 50; fax: +41 21 314 8299.
E-mail address: Francis.Verdun@chuv.ch (F.R. Verdun).
[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ejmp.2015.08.007

relevant image quality parameters and radiation dose indices must be defined, described, and used. This paper concentrates on image quality parameters.

The first step of the optimisation process should ensure that x-ray conversion into image information is performed as efficiently as possible. In projection radiology such as radiology or mammography one can use the DQE (Detective Quantum Efficiency as described in IEC 62220-1/2) as a global figure of merit. Unfortunately, due to the geometry and data processing required for CT, the use of such a quantity is not feasible. In general, one will assess the amount of radiation required to achieve a certain level of image quality. As a surrogate of the radiation received by the detector one uses the standardised CT dose index ($CTDI_{vol}$). This quantity represents the average dose delivered in PMMA phantoms of 16 and 32 cm in diameter and is related to the amount of noise present in an image. According to its definition, $CTDI_{vol}$ is different from the actual average dose delivered in a slice of a patient, and the latter should be estimated using the Size Specific Dose Estimator (SSDE) proposed by the AAPM (American Association of Physics in Medicine) [6]. For a given $CTDI_{vol}$ level, image quality parameters are generally assessed using the signal detection theory that considers the imaging system linear and shift invariant.

The next step of the optimisation process should be done with the clinical applications in mind. The determination of clinical performance is, however, difficult, expensive, and time-consuming. Furthermore, the results in these studies can be strongly dependent on the patient sample and on the radiologists involved. As an alternative, one can assess image quality using task-oriented image quality criteria. They will necessarily be simplistic in comparison to the clinical situations but make it possible to predict the perception of simple structures within an image. The phantoms available for this type of study remain quite simple but try to mimic important disease-related structures in actual patients. It is likely that 3D printing techniques will improve phantoms and task realism in the future [7–9]. To seek optimisation, task-based image quality metrics could be studied as a function of $CTDI_{vol}$ or SSDE. Figure 1 summarises this optimisation process.

Part 1 of this review focuses on signal detection theory and summarises the methods used to assess image quality in an objective way. When CT images are reconstructed using the standard filtered back-projection (FBP), these methods are commonly used to characterise a CT unit. The objective image quality metrics assess separate aspects of the features of the image, and therefore need to be combined to give an overall representation of the image quality.

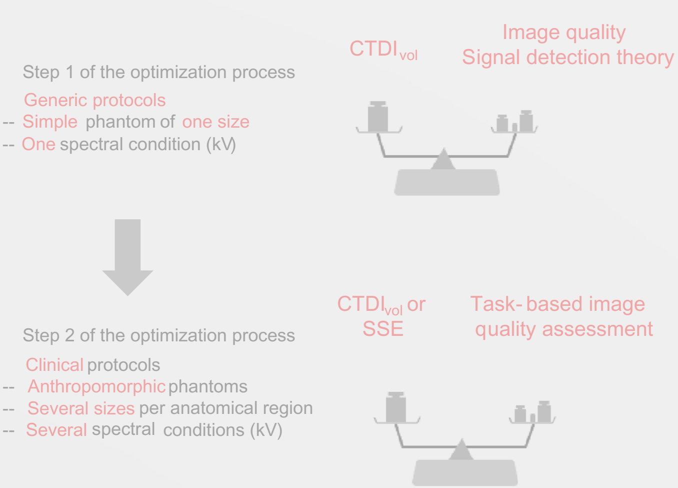

Figure 1. CT optimisation process in two steps: generic acquisition optimisation and clinical protocol optimisation.

To synthesise the information, and balance image quality with radiation doses, several figures of merit have been developed by combining image quality parameters such as the standard deviation in a region of interest (ROI) and the modulation transfer function (MTF). They were applied for specific clinical protocols to enable appropriate comparison of systems. This approach was quite useful during the development of CT technology, where performances between different units could vary drastically. These figures of merit can be based on simplified assumptions requiring caution in their interpretation. However it appears that the sensitivity of such methods is quite limited for newer systems, and, in addition, the effect of iterative reconstruction on the standard image quality parameters would mean that this approach would be difficult to implement.

Both clinical and phantom images can be assessed using the ROC paradigm or one of its derivatives (Localisation ROC, Free-response ROC). These methods give an accurate estimate of clinical image quality but, although carefully controlled measurements, they are still subjective because human observers are involved. These methods are time consuming and require large samples to obtain precise results. In spite of these limitations these methods can be used either by radiologists (when dealing with clinical images) or naïve observers when dealing with phantom images. To avoid the burden associated with ROC methods more simplified methods have been developed, for example, VGA (Visual Grading Analysis) in which image quality criteria can be used to give a relatively quick image quality assessment, without the explicit need for pathology or a task. Alternatively, phantom images can be assessed using the 2-AFC (two-alternative forced-choice) or M-AFC (multiple-alternative forced-choice) methods. Part 2 of this review discusses these methodologies, and these methods are used to validate the results produced by model observers presented in Part 3.

The introduction of iterative reconstruction in CT poses a new challenge in image quality assessment since most of the standard metrics presented in Part 1 cannot be used directly. In order to establish a bridge between radiologists and medical physicists, and therefore between clinical and physical image qualities, task related metrics can be used (even if the tasks are simplified versions of actual clinical tasks). Mathematical model observers are particularly suited to the routine image quality measurement of clinical protocols, with the results indicated to the user together with the standard dose report. Part 3 summarises the concepts behind these model observers, focusing on the anthropomorphic model observers that mimic human detection of simple targets in images, since the aim is to present them for practical applications. The theory and description of the ideal observer can be found in the literature and a brief introduction to this model is done at the beginning of Part 3. Note that model observers can also be used when images are reconstructed with FBP. The inconvenience associated with the use of model observers is that they lead to an overall outcome without the separation of the image quality parameters as with signal detection theory.

This paper is structured into three separate sections that provide an overview of the most common approaches taken when dealing with image quality in CT imaging. This structure is described in Fig. 2.

Traditional objective metrics

CT is a 3D imaging technique in which image quality assessment must be approached with some caution. Objective assessment of parameters that influence image quality is often made using physical metrics specified in either the spatial or spatial frequency domain. This duality is due to the fact that some features will produce overall responses which are independent of the location in the image, whereas other features will produce responses that are spatially correlated.

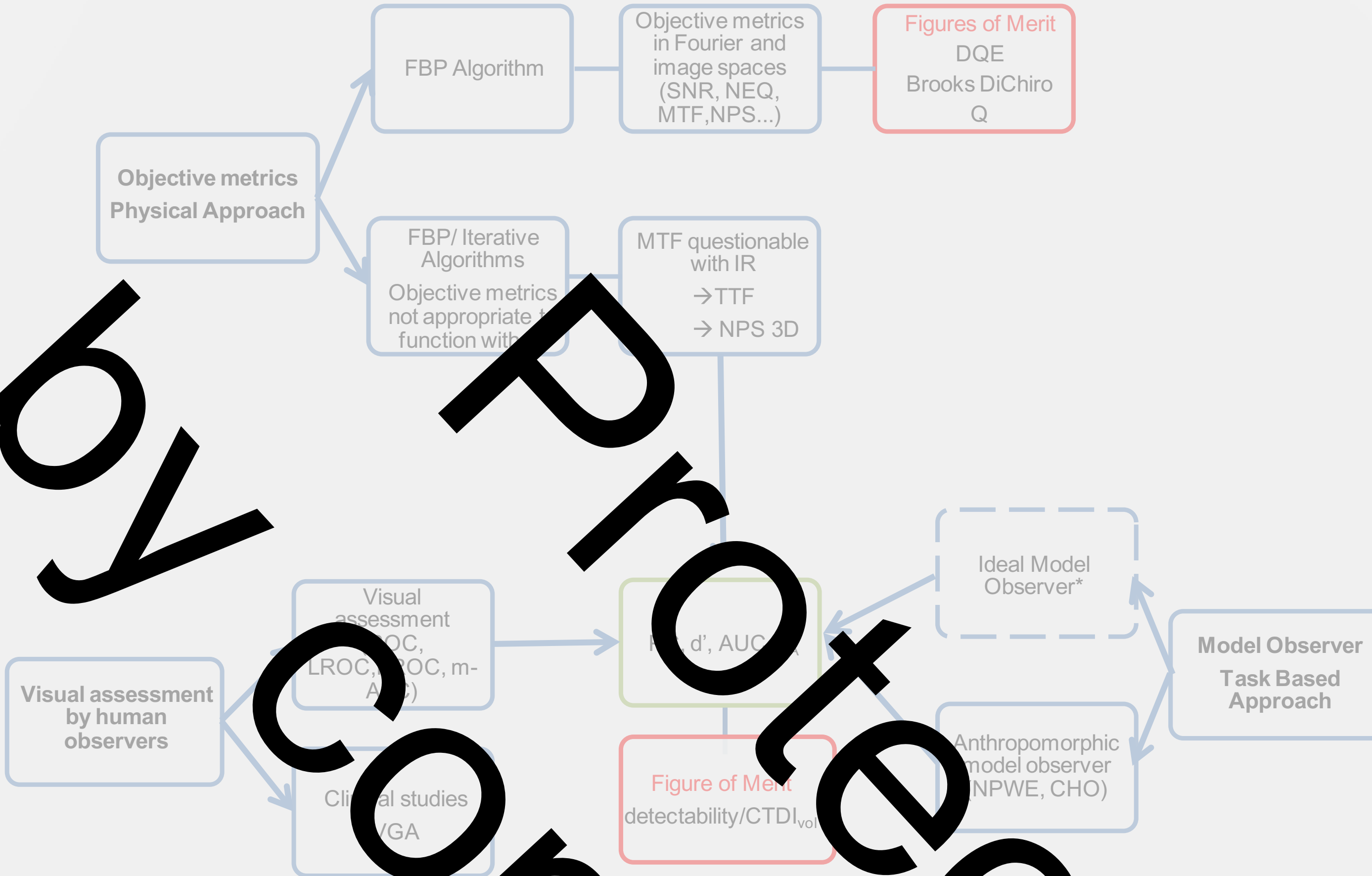

Figure 2. S[illegible]ary of the c[illegible]ent of the review (*this part will not b[illegible]resented).

Objective metrics in spatial domain

Image signal and image noise are key parameters [illegible] quality assessment. In the ideal and linear case, image signal (S) is directly linked to the detected number of photons N, whilst the noise (σ) may be seen as the pixel's stochastic fluctuation around their mean value. The photons are distributed according to Poisson's law, meaning that the quantity σ is equal to $\sqrt{N}$. The ratio of these two quantities yields the signal-to-noise ratio (SNR), expressed as:

$$SNR \propto \frac{S}{\sigma} = \frac{N}{\sqrt{N}} = \sqrt{N} \quad (1)$$

In an ideal device, each quantum could be counted by the detector and contributes towards the image. We could thus transpose Eq. (1) as:

$$SNR_{Ideal} \propto \frac{N_{Ideal}}{\sqrt{N_{Ideal}}} = \sqrt{N_{Ideal}} \quad (2)$$

However, due to the properties of the detector and its limited efficiency, a real measurement of the SNR would give the following result:

$$SNR_{Real} = \frac{N_{Real}}{\sqrt{N_{Real}}} = \sqrt{N_{Real}} < \sqrt{N_{Ideal}} \quad (3)$$

In Eq. (3), N_{Real} gives the number of quanta that contribute to the image for the real device and is also called noise-equivalent quanta (NEQ). Thus:

$$SNR^2_{Real} = N_{Real} = NEQ \quad (4)$$

[illegible]d on these parameters, we c[illegible] even[illegible] estimate the efficiency of a device by making the r[illegible] betw[illegible]n the number of photons [illegible]ually used for the imaging [illegible] number of ph[illegible]s to the detector. This quantity is called [illegible]ective qu[illegible] efficiency ([illegible]) and is defined as:

$$DQE = \frac{SNR^2_{Real}}{SNR^2_{Ideal}} = \frac{N_{Real}}{N_{Ideal}} \quad (5)$$

[illegible] Eq. (5), the NEQ [illegible] be [illegible]asured in a straightforward manner, but s[illegible]are m[illegible] be taken when estimating quantity SNR^2_{Ideal}. Indeed, when c[illegible]dering a m[illegible]ochrom[illegible]c beam, SNR^2_{Ideal} is simply the number of photons p[illegible]uced. [illegible]wever, for a polychromatic beam, SNR^2_{Ideal} should [illegible] sum [illegible]ariance of the number of photons in each energy bin. In[illegible]t, some authors prefer to use an energy weighted variance b[illegible]use most detectors integrate energy [10] to form an image.

Another commonly used global image quality index is the signal difference-to-noise ratio (SDNR), defined for an o[illegible]he intensity difference from the background divide[illegible] by the stan[illegible]rd deviation:

$$SDNR = \frac{I_{object} - I_{Background}}{\sigma} \quad (6)$$

These metrics are extended to the spatial freque[illegible]ain in the following section.

Objective metrics in Fourier domain

Spatial resolution can be defined as the ability to distinguish two separate objects and is directly linked to the pixel size, the reconstruction kernel as well as the hardware properties of the imaging device. In order to derive an expression for image resolution, it is necessary to describe the imaging process generating a CT slice. Our analysis will be restricted to the axial plane. $I(x, y)$, which is the image slice of an input object denoted by $f(x, y)$, can be mathematically expressed as:

$$I(x, y) = \iint f(x - x', y - y')PSF(x', y')dx'dy' \quad (7)$$

with $PSF(x, y)$ being the point spread function in the axial plane and describing resolution properties of the device. It corresponds to the impulse response of a system, the response of the system to a Dirac input $\delta(x, y)$.

Resolution can also be estimated through the line spread function (*LSF*), which is the response of the system to a straight line. Thus the relationship between the *LSF* and the *PSF* can be derived from Eq. (7) in which the input function is replaced by the equation of a straight line in the axial plane (that is to say replacing $f(x, y)$ by $\delta(x)$ in Eq. (7)), yielding:

$$LSF(x) = \iint \delta(x - x')PSF(x', y')dx'dy'$$

leading to:

$$LSF(x) = \int_{-\infty}^{+\infty} PSF(x, y)dy \quad (8)$$

The point spread function needs to be similar at each location in the image (shift invariance) in order to ensure that the *LSF* will remain the same at every localisation. However isotropy of the axial plane is a hypothesis which is not always true, especially when dealing with CT. In this case, the *LSF* will depend on the direction of the straight line in the axial plane. Assuming the straight line is positioned tilted with an angle θ the expression of the *LSF* will become:

$$LSF_\theta(x, y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} PSF(x', y')\delta((x - x')\cos\theta + (y - y')\sin\theta)dx'dy' \quad (9)$$

Besides those two metrics, it is also possible to estimate the resolution using the edge spread function (*ESF*), that is to say the response of the device to an edge. An edge can be mathematically approached by the Heaviside function $H(x, y) = \begin{cases} 1 & \text{if } x > 0 \\ 1/2 & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases}$. This function has the property: $\frac{dH(x)}{dx} = \delta(x)$.

Using this property, injecting $f(x, y) = H(x)$ in Eq. (7) and using Eq. (8) we obtain:

$$LSF(x) = \frac{\partial ESF(x)}{\partial x} \quad (10)$$

Hence, *PSF*, *LSF* and *ESF* are all related to each other and it is possible to use their representation in the frequency space thanks to the Fourier transform.

The Fourier representation of the *PSF* is the optical transfer function (*OTF*), which is defined as following:

$$OTF(u, v) \stackrel{\text{def}}{=} FT\{PSF(x, y)\} \quad (11)$$

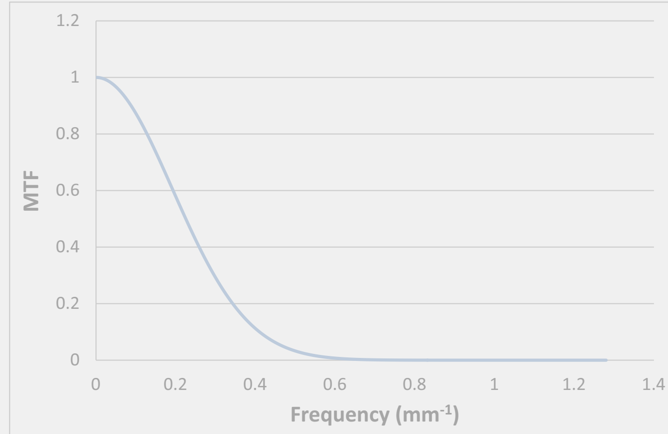

Figure 3. Example of a 1 dimension MTF curve of a GE VCT system with a 0.40 mm pixel size.

What is commonly used in order to estimate the resolution is the modulation transfer function (*MTF*), defined as the modulus of the *OTF normalised by its zero-frequency value*:

$$MTF(u, v) \stackrel{\text{def}}{=} \frac{|OTF(u, v)|}{|OTF(0, 0)|} \quad (12)$$

Using Eqs. (7), (11) and (12) together with the Fourier slice theorem and assuming shift-invariance in the axial plane, we can state that a normalised axial *MTF* of the system is given by:

$$MTF_{1D}(f) = \left| \frac{FT\{LSF(x)\}}{\int_{-\infty}^{+\infty} LSF(x)dx} \right| \quad (13)$$

This metric describes how well frequencies are transferred through the system and is therefore used to make objective resolution estimation (Fig. 3).

Practically, the MTF can be computed from the image of a point (~PSF), a line (~LSF) or an edge (~ESF) [11–13]. For calculating MTF from the image of a point source (effectively from the PSF), a metal bead or taut wire fixed within a dedicated phantom is used to generate the signal [14]. Boone [12] used a tilted aluminium foil of thickness 50 µm to generate an oversampled *LSF*; the MTF is then computed using Eq. (13). Judy [13] was the first to describe a calculation of MTF from an edge method in which the ESF was differentiated to give the LSF. The method has been developed over the years by various authors to include the use of spheres from which the oversampled ESF is built [15–17]. An older method was proposed by Droege and Morin, in which MTF is estimated from line pair test object images using the Coltman formula. Extensive details on the practical implementation of these techniques are given in ICRU Report 87 [18]. Several of these methods have been investigated by Miéville et al. in order to compare and contrast the advantages and drawbacks [19].

As with resolution, and of equal importance for SNR transfer, image noise can also be estimated in the frequency space. There are different sources of noise within the CT system, such as the electronic noise caused by the detector readout circuits (amplifiers) and the primary quantum noise which is inherent to the statistics of the limited quanta building the image. In a stationary system the Wiener spectrum or noise power spectrum (NPS) gives a complete description of the noise by providing its amplitude over the entire frequency range of the image [20]. If the image noise is not stationary, the Wiener spectrum is not a complete description and the full covariance matrix would be needed for complete description. However,

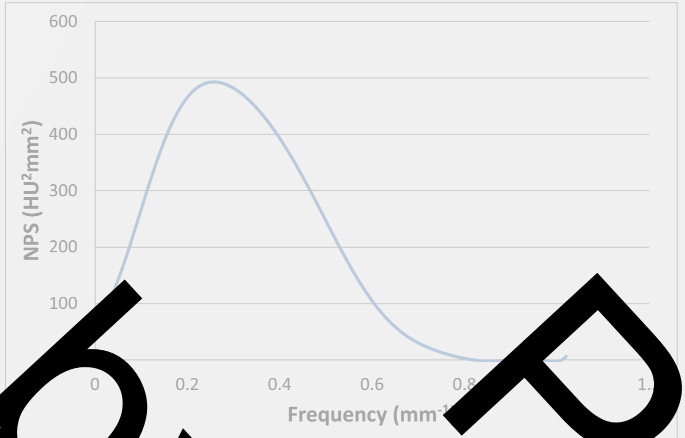

Fig. 4. Example of a radially averaged NPS obtained on a GE revolution system with a standard convolution kernel.

if applied with care – for example working with small ROIs, extracted from a restricted region of the image – the NPS can be applied to both conventionally (i.e. FBP based) and iteratively reconstructed images. For NPS calculation, the assumption of 'small signal linearity' has to be made in order to apply Fourier analysis, which requires system linearity in order to be valid. This is the case for the logarithmic step applied to all reconstruction processes and also to the explicitly non-linear iterative methods.

In order to compute the *NPS* of an image, it is necessary to acquire homogeneous CT images and select regions of interest (ROI) in this stack. The 2D *NPS* can then be computed as:

$$NPS_{2D}(f_x, f_y) = \frac{\Delta_x \Delta_y}{L_x L_y} \frac{1}{N_{ROI}} \sum_{i=1}^{N_{ROI}} \left| FT_{2D}\left\{ ROI_i(x, y) - \overline{ROI_i} \right\} \right|^2 \quad (14)$$

where Δ_x, Δ_y are the pixel sizes in the x and y dimension, L_x, L_y are the ROI's lengths (in pixel) for both dimensions, N_{ROI} is the number of ROIs used in the average operation and $\overline{ROI_i}$ is the mean pixel value of the ith ROI.

In practice, the *NPS* is largely affected by the detector dose, the hardware properties and the reconstruction kernel and algorithm. From each image of the stack a ROI is extracted and a custom computer program is generally used to compute the *NPS* according to Eq. (14). It is of common use to average the 2D *NPS* along a 1D radial frequency using the equation $f_r = \sqrt{f_x^2 + f_y^2}$ (Fig. 4). More details on the *NPS* computing can be found in ICRU Report 87 [18]. In the end, the NPS characterises the noise texture, thus giving a better and more complete description of noise than the simple pixel's standard deviation. Moreover, information about the pixel's standard deviation can still be retrieved with knowledge of the Wiener spectrum. Indeed, the Parseval theorem ensures that the total energy is obtained by summing the contribution of the different harmonics and that its value does not depend on the chosen space (image or frequency space). Since the *NPS* is a spectral decomposition of noise over frequencies, we have:

$$\sigma^2 = \iint NPS_{2D}(f_x, f_y)\, df_x\, df_y \quad (15)$$

As explained before, *MTF* shows how well the signal frequencies are transferred through an imaging system, that is to say it exhibits the signal response of a system at a given spatial frequency. As for the spatial domain, the ratio of signal (i.e. MTF) and noise (i.e. NPS) yields the output signal to noise ratio (the NEQ) and therefore the frequency dependent *NEQ* can be calculated as:

$$NEQ(f) = SNR_{Real}^2(f) = \frac{a^2 MTF_{1D}^2(f)}{NPS_{1D}(f)} \quad (16)$$

where a^2 is the mean pixel value squared.

The DQE in the frequency space can therefore be estimated by:

$$DQE(f) = \frac{SNR_{Real}^2(f)}{SNR_{Ideal}^2(f)} = \frac{a^2 MTF_{1D}^2(f)}{N_{Ideal} NPS_{1D}(f)} \quad (17)$$

Limitations of conventional and Fourier-based image quality metrics for the assessment of IR images

In order to compute an MTF that represents the spatial resolution of the entire image, the assumption of shift-invariance has to be made. That is to say that the device's response has to remain the same, whether measured at the image centre or periphery. If this assumption is not fulfilled it is necessary to make the measurements at the same location in different images to obtain an *MTF* that can be used to compare the resolution of different devices. Furthermore, the linearity hypothesis also needs to be fulfilled for the *MTF* to be reliable. That is to say, the output signal has to remain within the optimal range of response of the imaging system in terms of Hounsfield units (HU), usually in the range from −200 to +200 HU for clinical scanners [18]. Consequently, estimating the *MTF* with a high Z material can give a signal outside this range, yielding an incorrect assessment of resolution. In practice, estimating the *MTF* with high Z materials generally leads to a resolution overestimation because of the high SNR they generate [18].

Those two assumptions are approximately satisfied for CT images reconstructed with filtered back projection (FBP) algorithms and a standard reconstruction kernel, but the introduction of iterative reconstruction (IR) has changed the game [21]. Indeed, IR images exhibit stronger non-linear and non-stationary properties that force a change in the MTF measurement paradigm. Several authors have already highlighted the non-linearity problem of these algorithms, which manifests itself as contrast dependency of the resolution [21–23]. Also, investigations on how Fourier-based metrics are influenced by the characteristics of IR images have been done [24,25]. They showed, for example, that the shape of the NPS for some IR algorithms also depends on the dose level and that the resolution not only depends on the contrast but also on the radiation dose levels. These elements have highlighted the need to adapt the classical metrics to IR algorithms.

Adaption of Fourier metrics

These difficulties in estimating resolution can be overcome by using an adapted metric, such as the target transfer function (*TTF*), which makes it possible to characterise the resolution even in the presence of noise and contrast dependency [24,26]. *MTF* and *TTF* are similar but differ from one another in the sense that *MTF* only applies to a single given contrast level whereas a *TTF* will exhibit three different curves at three different contrasts (corresponding to three different materials) for the measurement (Fig. 5). This enables a characterisation of the resolution when dealing with non linear algorithms for which contrast influences the resolution. As already demonstrated by several authors this will make full characterisation of the resolution possible when dealing with IR [24,27].

The technological evolution of CTs has also led to changes in the way *NPS* must be computed. The 2D axial *NPS* was well suited for the first generations of devices where only one CT image per axial scan could be acquired without noise correlation between slices. Now that the acquisitions are also made in helical mode and that the number of detectors along the z-axis is higher, a 3D NPS is required to fully characterise the noise (Fig. 6) [12,28]. 3D NPS can

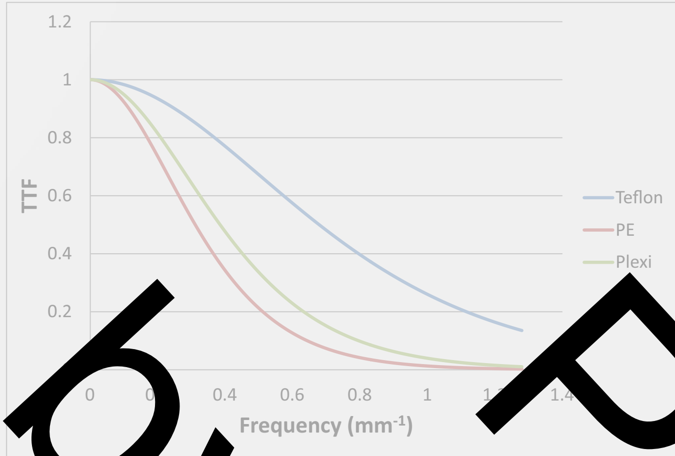

Figure 5. Resolution estimation through the *TTF* on a GE HD 750 system with 0.4 mm pixel size and three different materials (Teflon polyethylene and plexiglas). Differences are observed in resolution depending on the material. Such changes could not be observed when using the *MTF*.

be measured in a similar manner to the 2D NPS, but working with volumes of interests (VOI) instead of ROIs:

$$NPS_{3D}(f_x, f_y, f_z) = \frac{\Delta_x \Delta_y \Delta_z}{L_x L_y L_z} \frac{1}{N_{VOI}} \sum_{i=1}^{N_{VOI}} \left| FT_{3D}\left\{ VOI_i(x,y,z) - \overline{VOI_i} \right\} \right|^2 \quad (18)$$

For this case, the units of NPS are HU^2mm^3.

In this particular paradigm, Eq. (15) becomes:

$$\sigma^2 = \iiint NPS_{3D}(f_x, f_y, f_z) df_x df_y df_z \quad (19)$$

How to synthesise the information towards a figure of merit

Combining image quality and dose

In the clinical setting the focus for optimisation is balancing image quality and radiation dose in the context of the clinical question. Statistical noise, spatial resolution and imaged slice width are the fundamental parameters which describe the amount of object information retrievable from an image, and give rise to the perceived image quality. X-ray dose can be regarded as the cost of this information. It is meaningless to quote any of these image quality quantities without reference to the others, or to the radiation cost. The 'holy grail' is to try to find a way to combine the relevant parameters objectively and appropriately in a dose efficiency factor.

A dose efficiency factor, or figure of merit, can take a number of forms depending on how the various parameters are measured and quoted. Correctly developed and applied it can be used as a tool to compare scanner models, or simply different scan settings to optimise the balance of image quality and radiation dose.

How these parameters, resolution and noise in particular, are balanced is dependent on the clinical question and examination type. An important aspect that must be addressed is the influence of scan and protocol parameters that can be adjusted by the operator and how they affect image quality and radiation dose performance.

Clinical scanner settings – scan and protocol parameters

Any consideration of a theoretical approach to investigate a dose efficiency value needs to be in the scenario of the clinical question and the parameters used to create the image (Fig. 1).

Image quality and dose can be affected by the scanner design and also by the scan settings in the selected protocol (Table 1). The effect of the scan parameters, which form the examination protocol for the clinical question, can be seen in Table 2.

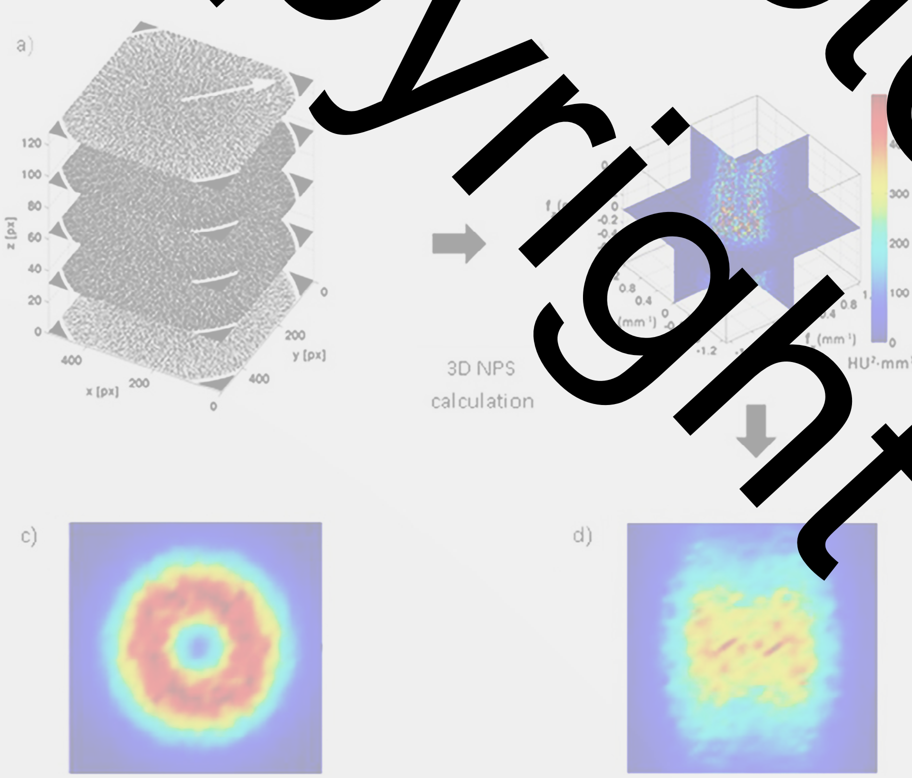

Figure 6. (a) The 3D homogeneous volume from which the 3D NPS is extracted. (b) The 3D NPS and the NPS sectioned in the (c) x–y (axial) and (d) the x–z (sagittal) planes. Figures extracted from Reference 22.

Table 1
Scanner design and scanner settings which can affect image quality and dose on scanner settings (courtesy of ImPACT [29]).

Scanner design factors	Scan protocol factors
Detectors material	Clinical application
Detector configuration	Tube current, tube voltage, focal spot size
Numbers of detectors, rows	Image reconstruction algorithms
Data acquisition rates	X ray Collimation width, detector acquisition width
Software corrections	Reconstructed image slice thickness
Filtration	Helical pitch
Focal spot size	Interpolation algorithms
Geometry (i.e. focus–axis, focus–detector distances)	

Combining image quality and dose metrics – theoretical background

The basic starting premise for a figure of merit for a dose efficiency parameter is that a dose efficient scanner will produce good resolution at minimum dose and noise.

There are a number of mathematical relationships that can be found in the scientific literature, both in terms of general imaging theory and CT in particular [30–32]. The two of interest for CT are Brooks and Di Chiro [33] and Riederer et al. [34]. These were used in the development of the ImPACT Q value which became a useful, and relatively widely known approach for comparing CT systems in the 1990s [35,36]. It was also explored by Fuchs and Kalender [37], more recently Kalender devoted a section to this subject in his book Computed Tomography: Fundamentals, System Technology, Image Quality, Applications [37,38]. However the fundamental relationship can also be found in standard textbooks on imaging with radiation [39,40]. The core of all these approaches is that the noise squared is inversely proportional to dose, and also inversely (in real or image space) proportional to the spatial resolution to the power 4. This encompasses spatial resolution in the x,y (to power 3) plane and also the z plane, quoted either as a size or frequency. In some equations the resolution is separated out into frequency for the x and y plane resolution, and the image thickness for the z-axis (z,x and z,y planes).

The relationship can be explored in more detail using the Brooks and Di Chiro equation [33]:

$$\sigma^2(\mu) = \frac{\pi^2 \beta \gamma(E) e^{\alpha} \mu_{en} E}{1200 \omega^3 h D} \tag{20}$$

Here σ^2 is the statistical error in the reconstructed image (i.e. the image noise); β is a beam spreading factor (non-parallel rays), $\gamma(E)$ is the average depth dose factor for photon energy (E), e^{α} is the logarithmic attenuation, μ_{en} is the energy absorption coefficient, E is the photon energy, ω is the detector aperture, h is the slice width, and D is the radiation dose.

For the purposes in this chapter, this can be simplified to:

$$\sigma^2 \propto \frac{1}{\omega^3 h D} \tag{21}$$

Similarly the Reiderer, Pelc and Chesler relationship is given as [33]:

$$\sigma^2 = \frac{\pi}{m N_p} \int_0^{2\pi} d\phi \int_0^{\infty} k dk \frac{|G(k)|^2}{k} \tag{22}$$

where m is the number of projections, N_p is the number of photons per projection, and $G(k)$ is the convolution function with frequency. The product mN_p could be regarded as a measure of radiation dose.

This essentially becomes $\sigma^2 \propto k_c{}^3/mN_p$ (where k_c is the cut-off frequency, i.e. the limiting resolution). Or, indeed as the paper states; 'for all valid correction filters ... σ^2 varies with the cube of the resolution'.

This is, in effect, the relationship of:

$$\sigma^2 \propto \frac{1}{N} \quad \text{i.e.} \quad \sigma^2 \propto \frac{1}{D} \tag{23}$$

where N is the number of photons and D is a measure of radiation dose for a fixed value of tube voltage. This can also be seen as a direct result from (1), assuming Poisson noise and without additive electronic noise.

Combining image quality and dose metrics – a practical approach

The discussion that follows is the approach taken by the UK CT scanner evaluation facility ImPACT [36]. It is a pragmatic solution to a complicated scenario of practical and computational effects on resultant image quality and dose for the operational CT scanner. This approach was reasonably successful for a number of years, and many scanner comparison reports were produced using this factor [32]. There is no other known work in this area covering a number of decades of scanner development. All measurements were undertaken according to a procedure with strict criteria, and in consultation with manufacturers as to the nature of their scan protocols, scanner features and reconstruction parameters. Measurements and analysis were carried out using typical clinical protocols, using the same image quality and dose assessment and calculation methods, and the same team of people. As scanners developed it became harder to apply such strict criteria, and with the development of adaptive filtration, and iterative reconstruction methods, it became very difficult to minimise the effects of other variables on ascertaining a dose efficiency value for a typical scan protocol.

$$Q \propto \sqrt{\frac{f^3}{[illegible]}} \tag{24}$$

Table 2
Dependence of image quality and dose parameters on scanner settings (courtesy of ImPACT, adapted from Reference 29).

	Noise	Slice width	Scan plane resolution	Dose
kV				
Effective mA (mA/pitch)				
Focal spot selection				
Pitch				
X-ray beam collimation				
Detector configuration (e.g. 16 × 1.25 versus 32 × 0.62)				
Scan time (for a given mAs)				
Interpolation algorithm				
Convolution kernel				
Reconstructed slice thickness				
Use of iterative reconstruction				

The dark blue represents a major dependence of image quality and dose on scanner settings and the soft blue represents a minor dependence.

where σ is the image noise, f is a measure of the in-plane spatial resolution (in frequency space), 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 facility [32,36] and first proposed in 1978 by Atkinson [35]. Initially one form of the generic equation was used, and then altered some of the definitions of the parameters involved, to create what became known as Q2 [31,41] as shown in Eq. (25).

The Q-factor (Q_2 factor) is in part empirical, it was used with caution and with strict adherence to the calculation procedure, which included standardising certain scan and protocol variables[illegible] it is not an absolute figure, it cannot be applied to the overall scanner, only to the examination protocol. Each set of image quality and dose parameters was therefore focussed on a typical clinical type of examination; for example a standard brain or standard abdomen. The first step in the process was to ascertain this scan protocol in conjunction with the manufacturer. Consideration of the effects of the scanner settings, as shown in Table 1, required some adjustment of the protocol. This was in order to minimise the effect of scan parameters[illegible] effects confounded the aim of comparison of image quality and dose, in the context of dose efficiency of the system. The associated challenge was to maintain the integrity of the suggested protocol for that type of examination. The second step was to undertake the various image quality and dose measurements and calculations, and then finally to apply the Q2 relationship.

$$Q_2 = \sqrt{\frac{f_{av}^3}{\sigma^2 z_1 CTDI_{vol}}} \quad (25)$$

The specific parameters used in calculating this value were measured using standard techniques and quality parameters such as would be used for quality control or acceptance testing:

σ = the image noise, the standard deviation of the CT numbers of a specified sized region of interest (5 cm[illegible]), expressed as a percentage (for water, standard deviation in HU divided by 10) measured at the centre of the field of view in a standard water phantom.
f_{av} = spatial resolution, given as $(MTF_{50} + MTF_{10})/2$, where MTF_{50} and MTF_{10} are the spatial frequencies corresponding to the 50% and 10% modulation transfer function values respectively (in line pairs per cm).
z_1 = the full width at half maximum (FWHM), (mm), of the imaged slice profile (z-sensitivity). This is measured using the inclined high contrast plates method (mm).
$CTDI_{vol}$ = volume weighted CT dose index (mGy).

To understand the dose efficiency relationship further in a practical manner, it can be helpful to consider the basic equation (Eq. 24) to be formed of three components:

$$\sigma^2 \propto \frac{1}{D}, \quad \sigma^2 \propto \frac{1}{z} \quad \text{and} \quad \sigma^2 \propto f^3 \quad (26)$$

which, in the Q_2 relationship, translate to:

$$\sigma^2 \propto \frac{1}{CTDI_{vol}}, \quad \sigma^2 \propto \frac{1}{z_1} \quad \text{and} \quad \sigma^2 \propto f_{av}{}^3 \quad (27)$$

Each of these relationships will be addressed more fully in the following sub-sections.

Dose value. The dose value in an earlier formulation of Q was the surface dose to a phantom, measured using thermoluminescent dosimeters. This was changed for Q_2 with the introduction of the standardised $CTDI_{vol}$ parameter. The cross-sectional averaging that contributes to the creation of the $CTDI_{vol}$ is more representative of the overall dose to the phantom and therefore a more appropriate value to be used.

The inverse relationship of dose with σ^2, $\left(\sigma^2 \propto \frac{1}{CTDI_{vol}}\right)$ has to be carefully considered with multi-slice CT beams. In CT it is generally acknowledged that the $CTDI_{vol}$ is a suitable dosimetry parameter; however the proportionality breaks down in MSCT since the penumbra contribution to the beam width is a constant value, and as such is a factor that affects the relative dose, and is not accounted for in the relationship. Therefore to accommodate this, the beam width needs to be kept as a constant when comparing one scanner to another, or to take it into account separately with a beam width correction factor.

Image slice width (z_1) – z-axis resolution. The effect on noise from the thickness of the slice (z_1) is from the imaged, as opposed to the nominal, slice width, with a dependence on the inverse proportionality of photons contributing to the image. For testing purposes the full width at half maximum (FWHM) of the imaged slice profile is a suitable parameter to use. However this does not fully describe the imaged slice profile, in terms of the photon distribution contributing to the reconstructed image. For ease of application the FWHM is used, even though a fuller description of this sensitivity profile would be [illegible].

Spatial resolution (f_{av}). A similar approach is taken with the spatial resolution parameter. Rather than using a single value from the modulation transfer curve, a more complete description of the resolution takes into account the full function over all frequencies, and a resolution value based on the average of the 50% and the 10% values of the modulation transfer function is therefore used. These values, averaged, do not completely describe the spatial resolution function, however they are common values automatically extracted from MTF curves as part of a standard testing process, and together were deemed to provide a better indication of the compromise between high and low spatial resolutions, compared to only one of the parameters alone.

The derivation of the cubed relationship of noise with spatial resolution ($\sigma^2 \propto f_{av}{}^3$) relies on assumptions of the shape of the convolution filter used (for example in Brooks and Di Chiro [illegible], the convolution filter is a ramp filter). In this way comparisons between scanners are likely to be more reliable when comparing images reconstructed with similar convolution filters and, in particular, algorithms that best represent ramp filters. These are in general, the filtered back projection filters named for 'standard' applications providing reasonably low spatial resolution in order to preserve the contrast detectability of an image. Filters that are slightly smoothed or slightly enhanced would be considered as close; however those with strong smoothing or strong edge enhancing would not be suitable. Reconstruction filters with 'standard' spatial resolution values were therefore chosen to minimise the dependency of Q_2 upon non-ramp like reconstruction filters. Fortunately, or appropriately, these were also the algorithms usually used in the standard clinical protocols under investigation. This aspect of the Q_2 equation is a pragmatic solution for the complexity of modern reconstruction algorithms. The reconstruction filter with MTF_{50} and MTF_{10} values as close as possible to 3.4 lp/cm and 6[illegible] lp/cm was used.

When investigating the empirical relationship with actual reconstruction filters, which range from ramp-like standard filters with conventional apodisation functions, to edge-enhancing high spatial resolution filters, it was found that the relationship was closer to a power of 4 or 5 [29,42].

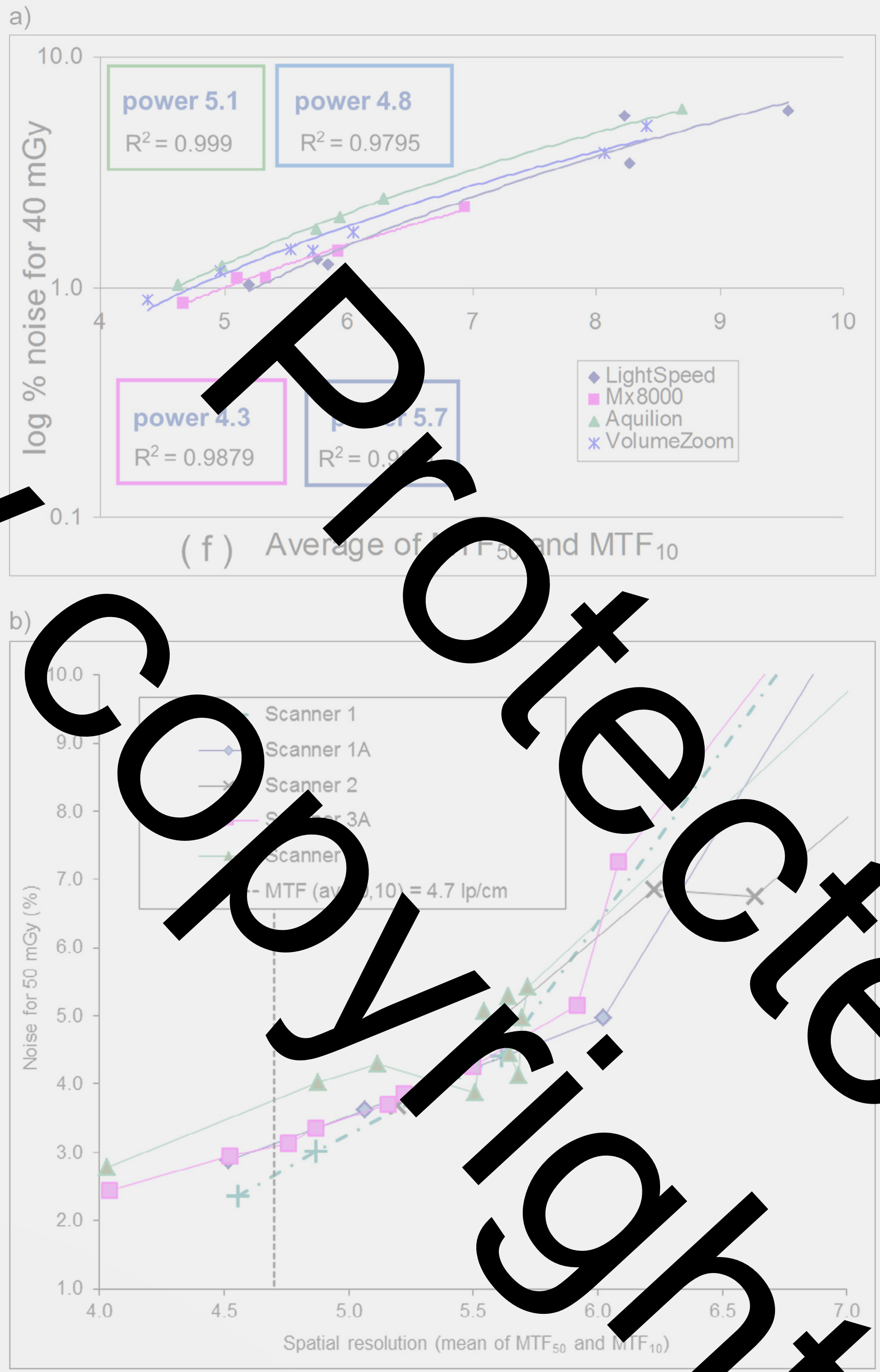

Figure 7. (a) Example for body algorithms, of logarithmic image noise against spatial resolution, with normalised dose (CTDI) demonstrating the deviation from the expected relationship. (The 'power' is the power to which f_{av} is raised against σ^2) (courtesy of ImPACT). (b) Head algorithms showing associated image noise against spatial resolution, with normalised dose (CTDI), demonstrating, particularly for scanner4, how small changes in spatial resolution give rise to large changes in measured noise [from data in Reference 41].

$$\sigma^2 \alpha \frac{f^{\approx 4-5}}{zD} \tag{28}$$

This is illustrated in the following graph (Fig. 7a), for the body scans. The different points on the graph relate to different reconstruction algorithms. This reinforces the need to compare the 'Q' for scanners with image quality parameters measured using standard algorithms only, as the cubed power relationship is not valid across the whole range of spatial resolutions.

However, with modern scanners and reconstruction algorithms, even with a 'standard' algorithm there can be anomalies in the expected relationships. With adaptive filtration and special reconstruction techniques, even selecting the lower spatial resolution algorithms, inconsistencies in the 'straight line' relationship can

Figure 8. Q2 values for several 16-slice scanners for standard head scans (courtesy of

appear, where a small increase in spatial resolution may not bring the expected associated increase in image noise, as shown in Fig. 7b [41,43].

The uncertainty in the Q value is estimated to be about 15%, and therefore, even once the confounding variables are standardised, it cannot be used to look for fine differences in the image quality and dose relationship [36,41,42], as shown for a set of 16-slice scanners in Fig. 8 [41].

However, it can demonstrate larger differences, such as with the difference between the dose efficiency of xenon gas and solid state detectors. Figure 9 shows data from the original 'Q' value, where surface phantom dose measurements were used (giving the surface multiple scan average dose (MSAD)). By normalising for the spatial resolution both in the z-axis (the image slice thickness) and the scan plane, this can be shown graphically as a relative dose.

Alternative method for combining parameters. Another approach to define CT dose efficiency was suggested by Nagel [44]. This approach for image quality determination is based on a statistical method of determining low contrast detectability (LCD) as previously suggested by Chao et al. [45]. In this method, a uniform phantom is scanned with specified dose and parameter settings. An array of square regions of interest (ROIs) is defined on the uniform image that is covering approximately a third of the central image area. By measuring the distribution of mean CT numbers of the ROIs and assuming a normal distribution, a prediction can be made of the CT number threshold of a low contrast detail having the same size as the ROIs in order to detect it at a 95% confidence level. This threshold contrast C is 3.29 times the standard deviation σ. This parameter is obtained by measuring the mean CT numbers of the ROIs before calculating their standard deviations. There is a 95% probability that a low contrast object of the same size as the ROIs is missed if the contrast is within the normal variation in the ROI means, i.e. if $C < 3.29\ \sigma$. Similarly, with a probability of 5%, a randomly high fluctuation of some ROI numbers could be mistaken for an actual low contrast object if the contrast of interest is sufficiently small. According to the Nyquist theorem, the ROI size limits the noise power spectrum (NPS) at a relatively low spatial frequency (here, approximately 1 lp/cm). Therefore, a measure of the detectability of low contrast objects having the same size as the ROIs suppresses spectral noise components at high spatial frequencies that are strongly affected by the detector and reconstruction algorithm.

The CT dose efficiency value (CTDEV) puts all parameters that are relevant for the specification of LCD into a single number that is based on the fundamental theory of Rose [46]:

$$CTDEV = 1.5\ \frac{e^{0.207(D_{eq}-16)}}{d^2 C^2 h_{rec}\ CTDI_{Vol,H}} \qquad (29)$$

with the diameter d of the low contrast detail (here, $d = 5$ mm), the slice thickness h_{rec} (in mm), the volume CT dose index $CTDI_{vol,H}$ for the 16 cm head phantom, the PMMA-equivalent phantom diameter (in cm) D_{eq} and the detail contrast (in %, with 1% = 10 HU) $C = 3.29\ \sigma$.

The method of Chao et al. can be easily implemented by applying customary CT phantoms and reduces the variability in LCD visually specified by human observers in conventional image quality assessments [45]. Chao's method has been applied by two CT manufacturers for the assessment of low contrast specifications [47]. The result of the method, however, depends on the size of the predefined ROI, the location of the image slice within the cone beam, and the filter used for image reconstruction [48]. As with other figures of merit, such as the Q_2 value, to apply the CTDEV for CT benchmarking, certain features must be standardised in detail. These are the protocol parameter set, reconstruction filter, phantom and method used.

Measures of diagnostic performance

Visual grading analysis (VGA)

Complementing the physical measurements of image quality, the assessment by observers is a subjective way to evaluate the image quality. Several general principles apply to all subjective observer studies. Patients should be selected to have a wide range of body habitus, they should involve as many observers as possible, and they should cover the range of expected competencies in the field [49]. When these assumptions are verified, visual grading analysis (VGA) based on observer scorings can be used to assess image quality.

VGA provides two types of information [50]:

Firstly, this subjective analysis provides information on the acceptability of the appearance (i.e. image noise level) of clinical images and how the anatomical structures are visualised. For example the VGA grades the visibility of important structures for different noise levels, because the detectability of low contrast structures is affected by noise, decreasing as the noise level increases.

Secondly, the subjective evaluation provides a context to interpret the physical metrics (i.e. *MTF*, *NPS*). Human observer evaluation is subject to change depending on context

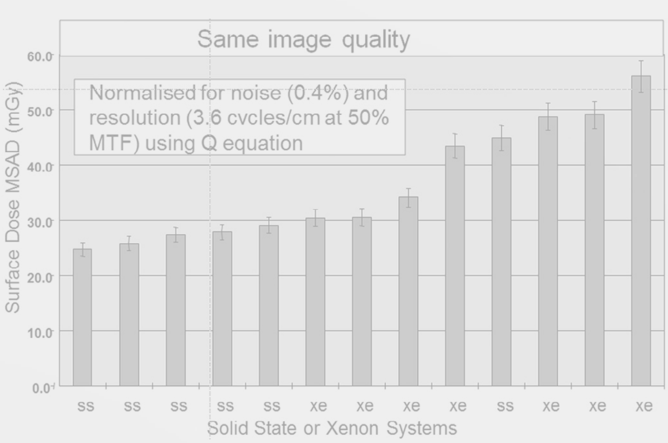

Figure 9. The use of a pervious version of Q to illustrate the relative dose, normalised by the other factors (courtesy of ImPACT) [slide 36 from Reference 29].

(i.e. brightness, tiredness), so the variability is not negligible and it is important to have a sufficient number of observers. For instance if a CT has 40% better *MTF* at high frequencies than another, but both CTs are rated by a single observer the difference between both systems will not become significant.

The VGA paradigm is split into two categories: relative grading and absolute grading.

Relative grading: The observer grades the image quality compared to a reference image or to the other images. The images should be displayed in random order to avoid any bias (i.e. first image read bias) and the viewing conditions should reproduce the planned environment of the reading diagnosis room [51]. The parameter studied should be as specific as possible, but it is possible to ask more than one question in order to evaluate several specific questions. The rating scale used in relative grading can have 3, 5 or more steps/ranks. The scale with 3 steps is not ideal because it is impossible to differentiate sufficiently. But when the degree of difference is small, a two step scale can be a possibility. The quality of the test is dependent on the reference image.

For instance a scale with 5 steps can be represented by:

–2: A is much better than B
–1: A is slightly better than B
0: A and B are equal
+1: B is slightly better than A
+2: B is much better than A

Absolute grading: The observers do not have any references and the images are displayed one by one. The evaluation is performed for one image at a time unlike the relative grading. To avoid bias from observer learning, the reading sessions must be separated in time. The grading scale should be numerical (i.e. from 1 to 10) or adjectival. With the adjectival scale, the descriptor should be expressive in order to create a difference between the worst and best cases. For instance, the Likert scale is a non comparative ordinal scale used especially in psychometric studies where the participants express their level of agreement with a given statement. Note that reproducibility is low with this type of study [52–54].

The results of a VGA study can be summarised with the VGA score (VGAS):

$$VGAS = \frac{\sum_{o,i} S_c}{N_i N_o} \quad (30)$$

where Sc = the given individual scores for observer (o) and image (i), Ni = total number of images, and No = total number of observers. In a VGA study to analyse the statistical difference, the analysis of variance (ANOVA) is calculated, associated with procedures for multiple comparisons.

For VGA, clinical images are required, which increases the implementation difficulties and also forces the avoidance of naïve observers. Indeed, to assess image quality in the VGA paradigm, the observer experience is very important if we want the obtained results to be as little distorted as possible. Nonetheless, VGA results are subjective and the analysis may be influenced by the experience of the radiologist, for instance in visualising different noise textures.

Decision theory: the statistical approach

It is common practice to specify the performance of diagnostic systems in physical terms as described in Part 1. However, it is complicated to translate these results to clinical performance. For instance, in detection tasks, certainty is rarely present. When an observer is asked to detect a signal on a medical image **g**, the result is a degree of belief that the signal is present. This degree of belief

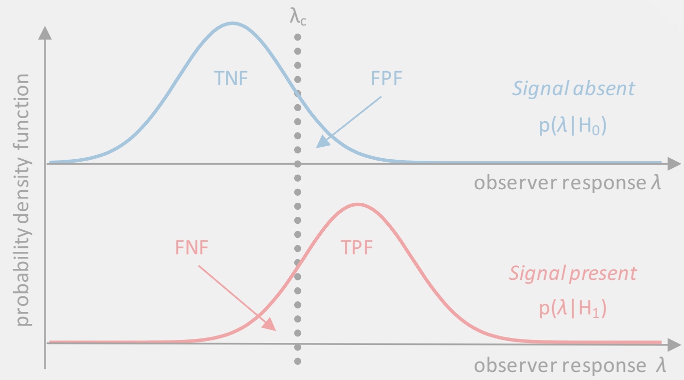

Figure 10. Probability density function of the observer response λ when presented with signal-absent images (top) or signal-present images (bottom). The vertical line λ_c indicates the threshold response above which the observer gives a positive response. TNF: true negative fraction; FPF: false positive fraction; FNF: false negative fraction; TPF: true positive fraction.

is commonly called the response $\lambda(\mathbf{g})$ of the observer: a low value denotes a confidence that the signal is absent, whereas a high value corresponds to the conviction that the signal is present. As shown in Fig. 10, the probability of obtaining a response can be plotted over all possible responses for two categories of images: those that do not contain a signal (top) and those that do contain a signal (bottom). These two curves are called probability density functions (pdf): respectively $P(\lambda|H_0)$ and $P(\lambda|H_1)$, where H_0 is the null-hypothesis corresponding to signal absent and H_1 is the alternative hypothesis corresponding to signal present. In radiology, the observer is forced to make a decision. In the present framework, this means that the observer chooses a threshold λ_c above which a positive decision is made. Below λ_c the observer makes a negative decision.

The integral of the distribution $P(\lambda|H_0)$ that is below the threshold is called the true negative fraction (TNF) or specificity. On the other hand, the integral of the distribution $P(\lambda|H_1)$ that is above the threshold is called the true positive fraction (TPF) or sensitivity. If the detection strategy is good, one expects both specificity and sensitivity to be as high as possible. However, Fig. 10 shows that changing the threshold changes the balance between specificity and sensitivity: increasing one parameter leads to a decrease of the other. There are mainly two ways to quantify the effectiveness of the strategy. The first is the signal to noise ratio defined as follows:

$$SNR_\lambda = \frac{|\mu_1 - \mu_0|}{\sqrt{\frac{1}{2}(\sigma_0^2 + \sigma_1^2)}} \quad (31)$$

where μ_0 and μ_1 are the means of $P(\lambda|H_0)$ and $P(\lambda|H_1)$, respectively, and σ_0 and σ_1 are the corresponding standard deviations. SNR_λ is a global figure of merit that broadly describes how two distributions are separated. This equation is similar to Eq. (6) about SDNR and its purpose is to compare two situations (with and without noise). However, Eq. (31) characterises the response of an observer and not a signal or a noise directly measurable on an image.

$SNR_\lambda = 0$ corresponds to the situation where the two pdfs have the same mean. If their shapes are the same, the decision based on such a strategy will be just guessing, and therefore the image does not transfer any information about the presence of the signal. A large SNR_t corresponds to well-separated pdfs. If the threshold is chosen between the distributions, then a large number of correct responses are expected.

A second way to quantify the effectiveness of the strategy is the receiver operating characteristics (ROC) curve, which shows all the possible combinations of sensitivity and specificity obtainable whilst

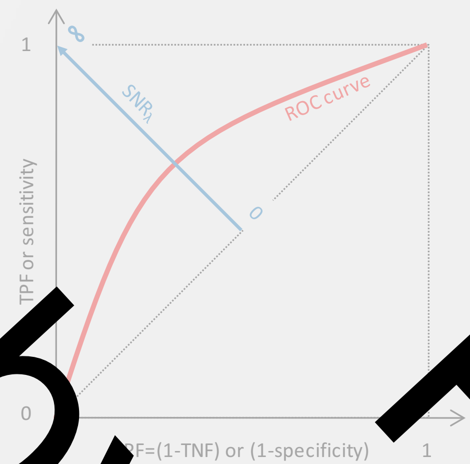

Figure 11. The ROC curve displays the true positive fraction versus the false positive fraction. If both response distributions are Gaussian with the same variance, then the intercept between the ROC curve and the second diagonal corresponds to SNR_λ.

we vary the threshold from the lowest to the highest possible values [55].

For historical reasons, the ROC curve displays the TPF versus the FPF, which is the sensitivity versus the (1 − specificity). If the pdfs are superimposed, the ROC curve is the straight line TPF = FPF. If the pdfs are well separated then the ROC curve has a square shape that passes close to the perfect point defined by sensitivity = 1 and specificity = 1. If pdfs are Gaussian with equal variances (this is often assumed in practice), the ROC curve is symmetrical and its intercept with the secondary diagonal corresponds to SNR_λ (Fig. 11). The value computed from the intercept between the ROC curve and SNR_λ is called the detectability index and usually represented with the symbol *d′*.

In practice, the observer (e.g. the radiologist) chooses a given threshold that corresponds to an operating point on the ROC curve. An objective way to define an optimal combination of sensitivity and specificity consists of computing the mean cost associated with all possible combinations of decision (negative or positive) and reality (signal absent or present):

$$\bar{C} = C_{00}P(D_0|H_0)P(H_0) + C_{01}P(D_0|H_1)P(H_1) + C_{10}P(D_1|H_0)P(H_0) + C_{11}P(D_1|H_1)P(H_1) \quad (32)$$

where C_{ij} is the cost associated with decision D_i and reality H_j, $P(D_i|H_j)$ is the pdf to make a decision D_i when the reality is H_j, and $P(H_1)$ is the probability to have a signal present. The latter is called prevalence in the case of the disease present in a population. By taking into account the basic properties of probabilities (e.g. $P(H_1) = 1 - P(H_0)$), Eq. (31) can be easily rewritten in terms of the four costs, sensitivity, specificity and prevalence.

All measures of clinical image quality using the decision theory are based on the truth. This truth can either be the ground truth (the truth is known exactly) or a gold standard (based on for instance the pathology outcome or experts opinion). Human observer studies are valuable as they are able to directly measure clinical image quality. Unfortunately, these methods are time consuming, expensive, and the inter- and intra-observer variability is often large. As a result assessment of clinical image quality is only applied incidentally. These limitations, together with the growing awareness of the importance of the evaluation of clinical image quality, make it more relevant to investigate whether model observers can be used as an objective alternative to human observers. This section is however limited to the discussion of rating scale experiments and m-AFC experiments using human observers. Part 3 provides an in-depth discussion about the use of model observers for this purpose. To gain insight into the decision making process rating scale experiments where observers are asked about their decision confidence can be performed. By varying variation in the decision threshold ROC curves can be drawn. The section "Rating Scale Experiments" provides more in-depth background of rating scale experiments. Another way to deal with observer decision criteria is by using multiple-alternative forced choice (m-AFC) experiments. In m-AFC experiments multiple alternatives are shown to the observer who is asked (forced) to choose the m-alternative which is most likely to contain the signal. This type of experiment will be discussed in detail in the section "Alternative Forced Choice Experiments."

Rating scale experiments

ROC analysis is a quantitative method applicable to a binary decision task. The method results in a graphical plot, the so-called ROC curve (Fig. 11), that illustrates the performance of observers (either humans or computer models) in the detection or classification tasks [50,56–58]. In this chapter we focus on the use of ROC analysis with respect to diagnostic imaging. In diagnostic imaging ROC studies, observers are asked to evaluate different cases and give a confidence level about the presence or absence of an abnormality in each case. The TPF and the FPF depend on the choice of the confidence level which results in a given decision (threshold). Generally, the ROC curve will be determined from the continuous confidence scale by varying the discrimination threshold. However, discrete binary confidence intervals can also be used in ROC analysis. An example of a continuous data experiment could be the assessment of the average CT number of pulmonary nodules from CT images to classify benign from malignant nodules (nodules with higher CT numbers are more likely to be calcified which is a sign of benignity; the average CT number will generate the continuous data). Discrete data could be obtained, for example, in a study with radiologists providing a five-point discrete confidence rating of abnormality concerning a set of normal and abnormal diagnostic images. For examples of ROC analysis used in computed tomography see references [59–61].

Theoretically, ROC curves are continuous and smooth. Unfortunately, the empirically derived ROC curves are most often jagged. Fitting algorithms can aim to create the smoothest curve according to the available data points. A wide range of algorithms is available for this purpose [56]. Often the area under the ROC curve (AUC or A_z) is determined as figure of merit for ROC studies. This AUC provides a summary measure of the accuracy of the diagnostic test that is independent of disease prevalence (in contrast to accuracy measures mentioned earlier). The AUC would be 1.0 for a perfectly performing test. A test performance that is equal to chance results in an AUC value of 0.5. Sometimes it can be more useful to look at a specific region of the ROC curve rather than at the whole curve. In these scenarios, it is possible to compute partial AUC. For example, one could focus on the region of the curve with a low false positive rate, which could be relevant for population screening tests [56]. The detectability, d_A, related to a rating scale experiment can be derived from the AUC:

$$d_A = \sqrt{2}\Phi^{-1}(AUC) \quad (33)$$

where, $\Phi = \int_{-\infty}^{x} \phi(y)dy$ is the cumulative Gaussian function and $\phi = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ is a Gaussian function.

If the decision variable distribution is Gaussian under both hypotheses (signal present and signal absent), and their variances are equal, then d_A is equivalent to d′.

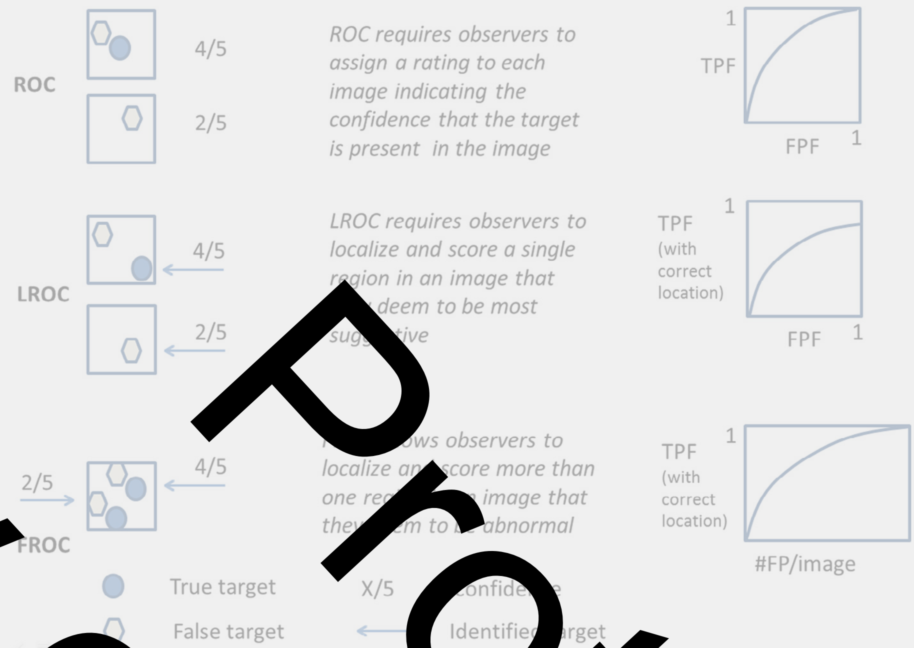

Figure 12. Related methodologies: ROC, LROC, FROC. The task in each of the methods is to give a confidence level concerning the presence of a true target (ROC) eventually in combination with the perceived location (LROC/FROC). In these examples the confidence level scale runs from 1 to 5. A rating of 4 on this scale is given as 4/5 (4 out of 5). Arrows indicate the perceived location.

Several advantages of ROC analysis can be considered. Among these is for instance the fact that the ROC approach provides a simple graphical plot that facilitates visual interpretation of data. Furthermore, depending on the implications of false positive and false negative results, and the prevalence of the condition, one can choose the optimal cut-off for a test from this graph, as the method provides a description of diagnostic accuracy for the full range of sensitivity and specificity. Moreover, two or more tests (for instance radiologists and a Computer Aided Diagnosis (CAD) system) can be compared, for example, analysing the area under each curve (where the better test has the largest AUC) [62]. Shortcomings of ROC analysis are related to its need for specialised computer software (regarding the curve fitting, AUC value calculation and confidence analysis on the ROC curve). Also, large sample sizes may be needed to generate reliable ROC curves. Finally, the ROC methodology does not optimally take the localisation task or the option of multiple abnormalities into account. For this purpose the so-called localisation ROC (LROC) and free response ROC (FROC) have been introduced. Figure 12 gives a graphical impression of the different methods and their concepts. Figure 13 gives a decision tree that illustrates the application of the different methods.

In LROC studies the observers' task is to mark a single location of a suspicious region in each case with a confidence level regarding the observed suspiciousness [56,57,63]. If the marked region is "close enough" to the true abnormal location, the observers' mark is considered a correct localisation. The definition of closeness is not uniformly defined and changes from study to study. Images with no targets (controls, benign, or negative cases) are also scored by marking a "most suspicious" area in the image and by giving this suspicious area a rating (forced localisation choice). To create an LROC curve, the TPF of decisions with correct localisation versus the FPF are plotted. It should be noted that the LROC curve does not necessarily pass the point (1, 1). Unlike the ROC methodology, in LROC the TPF of decisions with correct localisation may well be less than 1.0 at FPF = 1.0 because of incorrect localisations. Similar to the ROC methodology, the area under the LROC curve is considered to be a figure of merit of task performance.

To account for both the localisation and detection of abnormalities in images containing an arbitrary number of them, the free-response ROC (FROC) methodology can be used [56,57,63]. If the localisation mark is within a tolerance range around the true location and the rating of this mark is above a threshold, then a TP is realised. Otherwise a FP decision occurs. The free-response ROC curves are plotted by plotting the TPF (y-axis) versus the number of FP detections per image (x-axis) [64,65].

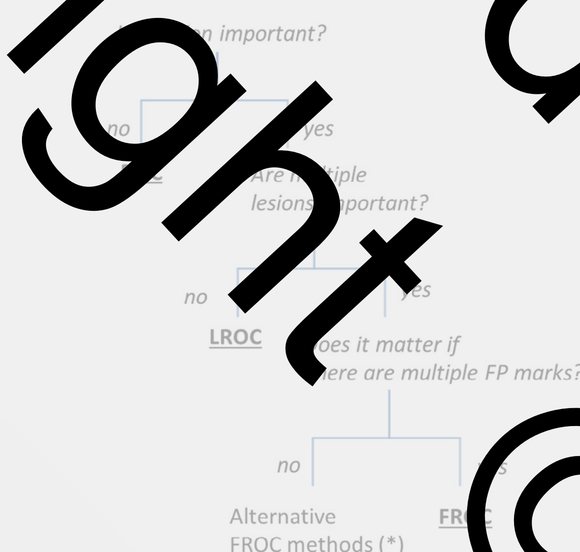

Figure 13. Decision tree illustrating the application of the different methods. The figure is a simplification of a figure provided by Wunderlich and Abbey [65]. Alternative methods (*) concern so-called Alternative FROC (AFROC) methods [54].

Alternative forced choice experiments

In forced choice experiments the observer has to make the decision 'signal present' between alternatives which are offered, even if this means that he has to guess. Compared to ROC studies, m-AFC experiments are faster and easier to perform [66]. However, m-AFC experiments do not provide insight into the underlying distribution functions and the trade-off between sensitivity and specificity [56]. Therefore, m-AFC is sometimes referred to as a poor measure of sensitivity [67].

The natural outcome of m-AFC experiments is a proportion of correct (PC) response. In m-AFC experiments and under assumption of Gaussian distribution of the decision variables [illegible], d′ and PC_m of [illegible] m-AFC task are related by:

$$PC_m = \int_{-\infty}^{\infty} \Phi^{m-1}(x)\phi(d' [illegible]$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $\Phi(x) = \int_{-\infty}^{x} \phi(y)dy$ are respectively Gaussian and cumulative Gaussian functions [68].

This [illegible] can be solved using tabulated values or numerical analysis (standard root finding methods) [69–72]. In the 2AFC experiment, this can be rewritten [illegible]

$$d' = \sqrt{2}\Phi^{-1}(PC_2) \quad (35)$$

For 2-AFC experiments, the PC is equal to the AUC but with human observers, the detectability obtained with the alternative forced-choice paradigm is larger than the detectability obtained with the ROC paradigm [50].

An example of setting for 2-AFC Signal Known Exactly/Background Known Exactly detection experiments is depicted in Fig. 14, where samples with signal present or absent are displayed together with a template of the target [illegible]

A detailed comparison and discussion about the use of R[illegible] and AFC experiments as well as the optimum selection [illegible] has been presented by Burgess [66]. This paper concludes that depending [illegible] the research question, a deliberate choice between ROC – m-AFC

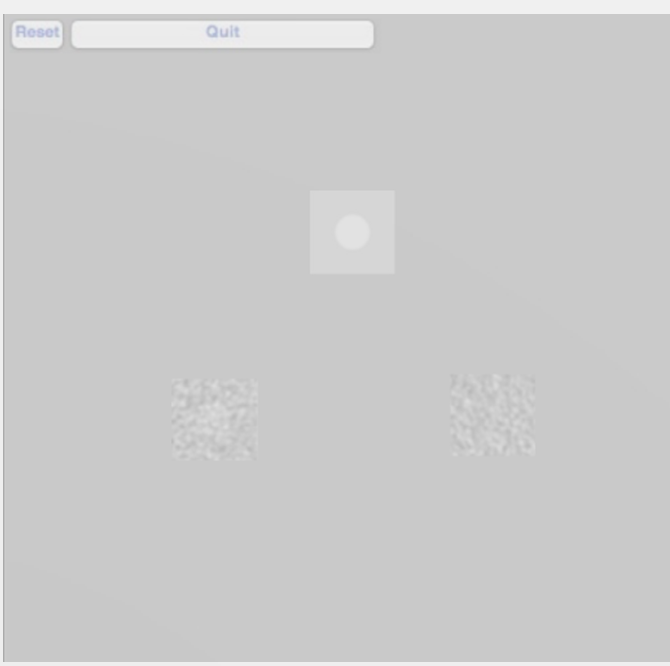

Figure 14. Interface of a 2-AFC human observer SKE/BKE detection experiment.

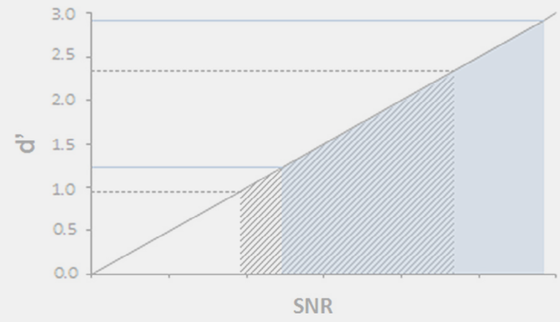

Figure 15. Selecting SNR range for a 2-AFC experiment (dotted black line) and a 4-AFC experiment (solid grey line).

experiments and the value of m is possible. In general m-AFC experiments are chosen if the study goal is to determine how well a certain task can be performed and when there is full control over both the ground truth and the SNR associated with the task. Most [illegible] m has a value of 2 or 4 but any scalar number larger than two is possible [73]. Burgess has demonstrated that a higher value of m will result in a smaller coefficient of variance. Besides this, he has shown that [illegible], for experiments with different values of m, is [illegible] against [illegible] signal to noise ratio (SNR) of the task they will fall on a th[illegible] line, independent of m [74]. From this it can be concluded that the [illegible] of m depends essentially on the SNR range for the experim[illegible] and [illegible] accuracy needed. The SNR range which can be used for an experiment is dictated by the SNR related to the lower thresho[illegible] (halfway between chance and 1) and 0.95–0.98. This upper level is advised to avoid issues due to observer inattention and their impact [illegible] [66[illegible]]. This means that in a 2-AFC experiment, the SNR range should be chosen to result in d′ values between 0.95 and 2.33, whilst this should be between [illegible] and 2.92 for 4-AFC (Fig. 15).

m-AFC experiments can [illegible] with m independent image combinations or single images which [illegible] divided into m areas in which the task can either be signal detection (present-absent) or classification (benign–malignant) [76,[illegible]]. [illegible] size for m-AFC experiments are based on the comparison of [illegible] expected difference between the PCs of the settings under evaluation for which standard statistical approaches can be [illegible]. m-AFC experiments are based on the signal-known-exactly (SKE) paradigm, which implies that clues should be provided regarding the signal and its position. Therefore, a template of the signal should be visualised together with the m alternatives and an indication of the possible location of the lesion should be indicated. [illegible] to provide clues on the signal position will result in non-linearity between SNR and observer d′ [66]. Finally, when designing m-AFC experiments care should be taken to avoid bias. For this purpose, the signal should be randomly assigned to one of the m alternatives and the observer PCs should be investigated for the tendencies to favor certain alternatives (e.g. the observer tends to choose left when he is unsure [illegible]).

Simulated and phantom images are generally well suited to conduct m-AFC experiments because of the full control of ground truth and SNR related to the task [66]. Phantom studies with the m-AFC paradigm are used to evaluate image quality of CT with both human and model observers [77–79]. But also for other modalities m-AFC methodologies are adapted into phantoms for quality control procedures like the CDMAM test object in mammography [80] or the CDRAD for general radiology [81,82].

Yes–no detectability experiments

In yes–no experiments observers only need to decide [illegible] the presence of an abnormality. Since yes–no experiments do not provide insight into the decision-making process of the observer they are not often used for measuring clinical performance very often. In the yes–no experiment the observer inspects one displayed image at a time and must indicate if the signal is present or absent. For a model observer, the yes–no performance is computed by comparing the decision variable to a threshold [50]. If the decision variable is higher than a threshold, the decision is: the signal is present. If the decision variable is less than the same threshold, the decision is: the signal is absent. In this test we assume that the case where the decision variable is equal to a threshold is negligible. With this performance it is possible to obtain four outcomes: true positive (the signal is present and the observer outcome is present), false positive (the signal is absent and the observer outcome is present), true negative (the signal is absent and the observer outcome is absent) and finally false negative (the signal is present and the observer outcome is absent). In the yes–no experiment the detectability index is given by

$$d_{YN} = \Phi^{-1}(TPF) - \Phi^{-1}(FPF) \quad (36)$$

The TPF represents the True Positive Fraction, and it means the probability given that the signal is truly present in the image. The False Positive Fraction represents the probability that when the signal is absent the observer indicates that the signal is present.

Model observers

ICRU Report 54 suggests that methodology based on statistical decision theory should be used in medical imaging [58]. Under this framework it is understood that the imaging performance depends on various factors: (1) measures describing the image contrast, image sharpness and the quantity characterising noise; (2) the detailed nature of the diagnostic task, including the clinically important details and the figure of the patient and the complexities arising from variability between patients; and (3) the degree to which information provided in the image is perceived by the clinician. Points (1) and (2) above are related to the information that is being recorded in the image data, but the ability of the human observer to extract the image information (Point 3 above) may also be an important or even the single limiting factor affecting diagnostic outcome.

Related to this, to simplify image quality assessment, the imaging process is often divided into two separate stages: the first stage consists of the image data acquisition and image formation stage; the second stage consists of the further processing of these data and their actual display to the human observer [58,83]. The first stage can be analysed rigorously by using the concept of the ideal observer, at least in principle and also in practice in simple cases. The ideal observer uses all available information in an optimal way for its decision; the performance of the ideal observer in a given imaging task can then be taken as a measure of the image information related to this task. The ability of the human observer to extract this image information can be measured separately; if the human observer is not able to use the recorded image information this implies leeway –and a need– to improve the image processing or display stage to be better suited to the human observer. This chapter will mainly concentrate on the imaging stage and leave the display stage largely outside the scope; the main aim of this paper is to review methods for evaluating CT scanners and their performance and not the quality of display equipment and display conditions. However, some methods which try to include features of human observers are shortly presented.

The performance of the ideal observer can usually be evaluated only for simplified classification tasks, such as the signal-known-exactly/background-known-exactly case, denoted as SKE/BKE. In this case the ideal observer has all a-priori information of the task, and its performance for classifying images to signal-present and signal-absent cases depends only on the amount of information in the image [58]. The performance of the ideal observer can therefore be taken as a measure of the task-related image information. Other tasks, involving uncertainty of the signal and the background, would be better related to clinical image quality assessment than the SKE/BKE. In such tasks the performance of the observer is not just dependent on the information in the image. The amount of a-priori information about the task that the observer has needs to be taken into account and will affect the performance. It may then sometimes be difficult to quantify the actual effect that this a-priori image information has in the task performance.

Relying on stylised imaging tasks based on the SKE/BKE paradigm may not always be reasonable; see, e.g., Myers et al., where the problem of aperture-size optimisation in emission imaging was considered and it was shown that the optimal aperture would be highly different for the detection of a simple signal in a known background and in a lumpy background [84]. Often, however, it may be considered plausible that the performance of an imaging system in tasks involving incomplete a-priori information could be monotonically related to the outcome in similar detection tasks in the case of full a-priori information (SKE/BKE) [85–88]. This appears to be the case in the paper of Brown et al., where the ideal observer's performance was studied for the signal position unknown case [89]. However, we are still far from completely understanding how a-priori information and the actual image information interact in medical imaging.

In phantom measurements the variability and non-uniformity of real patient images are not usually present. In the SKE/BKE paradigm any background structure is treated as being a deterministic known structure, which does not impair detail detectability. This may not always be realistic for a human observer, whose detection performance may in some cases be more impaired because of background variability than because of actual stochastic noise [90–94], but is certainly applicable to the ideal observer. Human observers seem to operate somewhere between two interpretations: background variability appears to function as a mixture of noise and deterministic masking components. For a more detailed discussion on this matter, see, e.g. Burgess and the references therein [91]. For a thorough presentation of modern image science, see the book by Barrett and Myers [57] and by Samei and Krupinski [56]. Another useful handbook on imaging systems, image quality and measurements has been published by the International Society for Optical Engineering [50]. Also, a discussion and review of task-based methods for assessing the quality of iteratively reconstructed CT images have been published recently [25]. They conclude that Fourier-based metrics of image quality are convenient and useful in many contexts, e.g., in quality assurance, but the assessment of iteratively reconstructed CT images requires more sophisticated methods, which do not rely on assumptions of system linearity and noise stationarity; these assumptions are prerequisites in the Fourier-based methods [95–97].

Linear observers

Mathematical theory

A linear observer can be described with a decision statistic $\lambda(g)$ which is a linear function of the image data, instead of being a more general function. In the vector notation of images this can be written as an inner product of a template $\mathbf{w}$ and the image $\mathbf{g}$:

$$\lambda(\mathbf{g}) = \mathbf{w}^{\mathrm{T}}\mathbf{g} \quad (37)$$

The non-zero elements of the template correspond to image locations where the pixel value needs to be taken into account, and by what weight. The weight can be either positive or negative. Pixels with the value zero in the template do not influence the decision statistic at all, and the observer considers the data in those pixels to be irrelevant for the decision.

The importance and frequent use of linear observers stems mainly from their manageability and ease of use. Further, as was seen in the preceding chapter, the ideal observer of many cases may be obtained in a linear form. This is not the case for all detectability tasks, however. For example, the ideal detection in the case involving uncertainty of the signal position will result in a non-linear test statistic (see, e.g., Brown et al. [89]). A linear observer for this task would consist just of a template which is obtained as the convolution of the pdf of the signal position and shape. Therefore, essentially, this observer would measure only the mean brightness of the image and it seems clear that it would be much less efficient than a human observer, for example.

In order to compute the SNR of a linear observer, we first need to express the linear response under hypothesis H_j as well as its associated variance:

$$\bar{\lambda}_j = \langle \lambda(\mathbf{g}) | H_j \rangle = \mathbf{w}^T \langle \mathbf{g} | H_j \rangle$$
$$\sigma_j^2 = \left\langle \left(\mathbf{w}^T\mathbf{g} - \langle \mathbf{w}^T\mathbf{g} | \mathbf{H_j} \rangle \right)^2 \middle| H_j \right\rangle = \mathbf{w}^T \mathbf{K}_j \mathbf{w} \tag{38}$$

This allows us to easily express the signal to noise ratio of a linear observer by injecting Eq. (38) into Eq. (3[illegible]):

$$\mathrm{SNR}_\lambda^2 = \frac{\left(\mathbf{w}^T(\langle \mathbf{g} | H_1 \rangle - \langle \mathbf{g} | H_0 \rangle)\right)^2}{\mathbf{w}^T \frac{1}{2} (\mathbf{K}_0 + \mathbf{K}_1) \mathbf{w}} \tag{39}$$

Here, it is important to recall the assumptions required for Eq. (39) to be meaningful. First, this requires that the conditional distributions of λ are normal. This is the case at least when the noise in the images is multivariate normal. Secondly, if the covariance matrices for the signal and background cases are different, the SNR does not define the entire ROC curve, but the area under the ROC curve and the percentage of correct answers in a two-alternative forced choice test using the same images are still specified by the SNR. An inequality of covariance matrices $\mathbf{K}_0$ and $\mathbf{K}_1$ would also infer that a linear observer is not ideal, and may fall far beyond the true ideal observer [98]; however, if measured covariance data are used it is useful to improve the precision of the **K**-estimate by including both measured covariance, $\mathbf{K_0}$ and $\mathbf{K_1}$.

By inserting the w-templates of the PWMF and the NPWMF to Eq. (39) we obtain the well-known expressions for their SNR

$$\mathrm{SNR}^2{}_{\mathrm{PWMF}} = \mathbf{s}^T \mathbf{K}^{-1} \mathbf{s} = \mathbf{S}^T \mathbf{W}^{-1} \mathbf{S} \tag{40}$$

and

$$\mathrm{SNR}^2{}_{\mathrm{NPWMF}} = \left(\mathbf{s}^T\mathbf{s}\right)^2 / \mathbf{s}^T\mathbf{K}\mathbf{s} == \left(\mathbf{S}^T\mathbf{S}\right)^2 / \mathbf{S}^T\mathbf{W}\mathbf{S} \tag{41}$$

where we have denoted the Fourier transform of **s** by **S** and that of matrix **K** by **W**. If the noise is stationary, **W** is a diagonal matrix and its diagonal values represent the NPS. Then, decomposing the SNR^2 to components: each frequency k contributes by amount

$$\mathrm{SNR}^2{}_{\mathrm{PWMF},k} = \mathrm{I}|S_k|^2 / W_k \tag{42}$$

to the total $\mathrm{SNR}^2{}_{\mathrm{PWMF}}$. This simplicity is lost if W is not diagonal.

The best possible linear observer is called the Hotelling observer. The Hotelling observer is equal to the PWMF in the case of signal-independent (additive), normally distributed noise and both of these reduce to the NPWMF, when the noise is white. As discussed above, the Hotelling observer may also fall far below ideal performance, for example, in the signal position unknown case, where the ideal decision statistic is not a linear function of image data [89].

The strategy of the ideal observer may be complicated by **K** not being diagonal. However, in the case of uncorrelated image noise the strategy is self-evident: the ideal observer then just looks more keenly to image pixels where the presence of the signal is known to have a strong effect and where the uncertainty of the measurement (noise) is small. Image areas that are not affected by signal presence need not be observed at all. This same interpretation applies to the case of coloured, stationary noise as well; then the Fourier transformed data will have a diagonal covariance matrix, where the diagonal elements constitute the noise power spectrum. In this case the ideal observer puts more emphasis on spatial frequencies where the signal presence makes a large contribution and less emphasis on frequencies which contain more noise.

If the image noise is not white, the NPWMF observer is suboptimal because it does not take into account the noise correlations between pixels, or equivalently, the different noise power at various spatial frequencies. Therefore, in this case, the observer is not tuned against the noise similarly as the ideal observer and it shows a penalty in its performance. However, if the frequency spectrum of the signal is concentrated on a relatively narrow band of frequencies where the frequency dependence of the NPS is modest, one can expect the NPWMF observer to perform nearly as well as the ideal PWMF does. This may happen, for example, when the signal to be detected does not have sharp details and is of a relatively large size.

However, note that by definition, the NPWMF believes that the background level is equal in all images and therefore needs not be observed. The NPWMF measures the image intensity only in the pixels that belong to the expected signal position and discards the data in all other pixels. For a disk signal this would be equivalent to observing just the total image intensity of the signal disk area and masking away all other image areas: no reference to the contrast between the signal and the background is made. If in fact, there is any – even small – variation in the background level from image to image, or if there is any low-frequency background variability (e.g., variable lumpy background structures) which in reality can have an effect on the image intensity in the signal detail area, the NPWMF can be considered as being a mislead observer, which will perform extremely poorly and often performs worse than human observers. This was the case, for example, in a paper that considered signal detection in added low-pass correlated noise and found that the NPWMF observer was very inefficient and even humans significantly outperformed it [99]. This and other similar results greatly diminished the interest in the NPWMF observer.

To improve this situation, Tapiovaara and Wagner [illegible] introduced the DC-suppressing observer, which leaves the average brightness of the image (or the zero-frequency channel) outside of the decision.[1] This observer is achieved by subtracting the mean pixel value of the PWMF template from every pixel of the template

$$\lambda_{\mathrm{DCS}} = \left[\mathbf{s} - \left(\mathrm{N}^{-1}\Sigma s_k\right)\mathbf{1}\right]^T \mathbf{g} \tag{43}$$

Here, N is the number of pixels in the analysed image area and **1** denotes a vector with all elements equal to unity. In the Fourier domain this observer is:

$$\lambda_{\mathrm{DCS}} = \left[\mathbf{S} - S_0\mathbf{e}^0\right]^T \mathbf{G} = \sum_{k=1}^{N-1} S_k^* G_k \tag{44}$$

[1] In practical imaging measurements one often does not analyse the whole image area, but considers only a relatively small sub-area containing the signal and a reasonable surround of it. Then the image vector **g** corresponds to this sub-area and the zero-frequency of this image data includes contributions from very low-frequencies in addition to the strict zero-frequency of the whole image data.

This modification of the NPWMF-observer turned out to be crucial for the performance of the observer in measurements of fluoroscopic imaging, where excess noise in the mean image brightness strongly and variably impaired the performance of the NPWMF [100]. This zero-frequency variability can be assumed to be common in other fields of radiology as well: the exact mean image brightness is not probably an important diagnostic feature in any imaging modality, and, on the other hand, if there is excess variability in the brightness, including it – as the NPWMF does – will result in a notable performance penalty. Such a variability in average brightness can be seen as a delta spike at the origin of the NPS and can be properly weighted by the PWMF, of course. However, in many recipes for measuring the NPS, the DC-component is normalised out and therefore equals zero in the NPS results (e.g. Boedeker et al. [101]). While noiseless data in real systems is not realistic, it is then important not to include the zero frequency signal component in the SNR calculation either.

Non-prewhitening with eye filter

Another modification of the NPWMF includes filtering of the image with an eye-filter, intended to obtain a better agreement of the performance of this model observer and human observers. The observer is often denoted as NPWE [102] (a similar observer model has been presented earlier by [illegible] et al. [103]). This observer is usually expressed in the spatial frequency domain and the eye filter **E** mimics the visual spatial frequency response function (or the contrast sensitivity function) of the human eye. The application of **E** requires knowing the dimension of the image and the viewing distance. The decision function of this observer is then:

$$\lambda_{NPWE} = [\mathbf{ES}]^T \mathbf{EG} = \mathbf{S}^T\mathbf{E}^T\mathbf{EG} \tag{45}$$

It is noted here that the eye filter also suppresses the zero-frequency, like the DCS-observer above, but the NPWE observer also takes very low frequencies into account with a low weighting. This is the main factor for the NPWE observer performing much better than the NPWMF in studies involving excess noise in very low frequencies [25,102]. This means that the usefulness of this observer model may actually be more related to its suppressing of low-frequency noise than in its attempt to mimic human vision.

As an example of NPWE performance, Fig. 16 shows the detectability index (d′) or SNR as a function of object diameter for the 0.5% contrast group of the Catphan and three mAs levels acquired in a Toshiba Aquilion ONE 320 detector-row CT scanner. The NPWE detectability improved with increasing mAs, as the noise level of the images decreased, for all the objects [50].

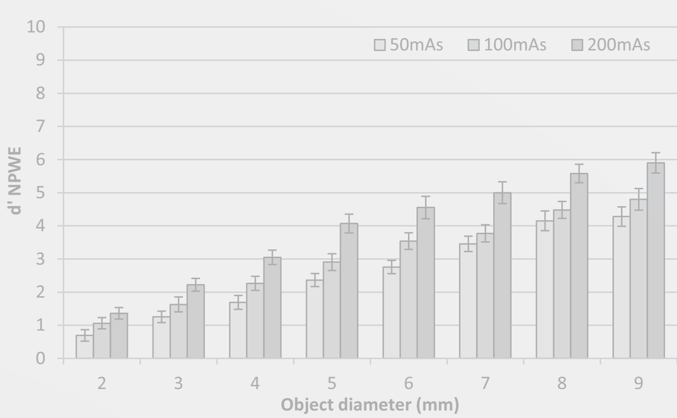

Figure 16. Detectability index (d′) as a function of object diameter for the different levels of mAs for the 0.5% contrast group (2-9 mm) in the Catphan 600 Phantom (Phantom Laboratories, New York).

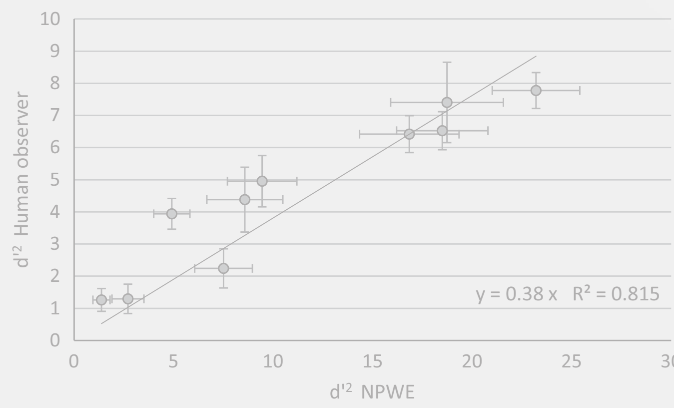

Figure 17. Detectability index (d′) of the average human observer as a function of the NPWE model observer d′, both squared, for 1% contrast objects and all dose values. The efficiency, η, tallies the slope of the linear fit.

The detectability index is given when two assumptions are verified [illegible]. Firstly the template responses must be Gaussian and secondly the template responses are statistically independent [90]. The performance is given in terms of distance in standard deviation units between the signal distribution and the noise distribution.

$$d' = \frac{\langle\lambda_s\rangle - \langle\lambda_n\rangle}{\sigma_\lambda} \tag{46}$$

where λ_s is the mean model response to the signal, and λ_n is the mean model response to the background. σ_λ is the standard deviation of the model response.

The advantage of this metric is that it computes directly from the image statistic.

Model observers can also be otherwise modified in order to mimic human performance better, for example, by including internal noise [104,105]. Internal noise degrades the model performance, and takes into account the fact that human observers have "noise" by not giving necessarily the same answer when a certain image is presented twice or more to be scored [illegible]. Many approaches can be used to decrease the model's performance, and each has pros and cons [illegible–107]. Such models are of interest in efforts to reproduce the efficiency of the visual detection performance of humans but are not expanded in this review. In Fig. 17 the PC values were translated into d′ and an efficiency (η) was calculated to normalise the model observer results, fitting d′ human as a function of d′NPWE, both squared. As the curve representing d′ as a function of PC saturates above PC ≈ 0.9[illegible] for 2-AFC experiments, only the values below this threshold were taken into account [108].

Channelised Hotelling Observer

Another type of linear observer models is the Channelised Hotelling Observer [109] (CHO) either with or without internal noise; only the latter model is considered here. A thorough treatment of both can be found in Abbey and Barrett [110]. The motivation for this observer results both from its effect in reducing the image data from a large number of pixels to a much lower number of scalars, called channel responses and by the ability of such models to mimic the detectability results of human observers. If the channels in the model are selected such that they help in the tuning against the noise frequencies without losing too much of the signal energy they may also provide an improvement over the non-prewhitening observer types and a useful approximation for the ideal observer. The reduction of dimensionality especially simplifies estimating and inversing the covariance matrix.

The CHO does not have access directly to the pixel values (or the Fourier transform) of the image. Instead, first the image data (**g**) are linearly combined to a small number of channelised data (**u**) by multiplication with a matrix **T**:

$$\mathbf{u} = \mathbf{T}^{\mathrm{T}}\mathbf{g} \tag{47}$$

Here the column vectors of **T** represent the spatial profiles of the channels. These channelised data are then combined with a weighting template **v** to a linear decision function:

$$\lambda = \mathbf{v}^{\mathrm{T}}\mathbf{u} \tag{48}$$

If the noise in the image data **g** is Gaussian, it is also Gaussian in the channels **u**, and we already know that the ideal observer (which, however, has access only to the channelised data) is obtained with weighting $\mathbf{v}^{\mathrm{T}} = (\mathbf{u}_1 - \mathbf{u}_0)^{\mathrm{T}}\mathbf{K}_{\mathbf{u}}^{-1}$, and the decision function of the observer is:

$$\lambda_{\mathrm{CHO,T}} = (\mathbf{u}_1 - \mathbf{u}_0)^{\mathrm{T}}\mathbf{K}_{\mathrm{u}}^{-1}\mathbf{u} = (\mathbf{u}_1 - \mathbf{u}_0)^{\mathrm{T}}\mathbf{K}_{\mathrm{u}}^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{g} \tag{49}$$

Above the channels were presented in the image domain. Usually, however, the channels are specified in the frequency domain, and may be either non-overlapping frequency intervals or overlapping functions of various forms, such as sparse or dense difference-of Gaussians, Laguerre–Gauss polynomials or other functions [109,111].

Note that in the case of stationary image noise the non-overlapping channel models result also in diagonal covariance matrix, because the frequency channels remain independent, whereas the overlapping channels cause correlations in the noise. If one prefers working in the image domain, one can obtain the spatial representations of the frequency selective channels by taking the inverse Fourier transforms of the latter.

In image quality assessment when using these channelised models it is important to note that the channelised Hotelling observer can adapt to the signal and the image covariance only after they have passed through **T**. Then, for example, the observer is sensitive only to signals that cause a change in the channelised signal $\mathbf{T}^{\mathrm{T}}\mathbf{s}$ (or, equivalently, in the frequency domain representation). For sparse channel models with just a few channels, a significant loss of information may occur in the formation of the channel responses [110].

Also, these observers are typically zero-frequency suppressing, although, being tuned against the noise in the different channels, they could also otherwise handle variability in the average image brightness better than the NPWMF. This would require, however, that if zero-frequency is included in the lowest frequency channel, not much of the important signal energy shall be included in this channel.

Usually, in applications related to medical imaging, the channels are defined to be cylindrically symmetric and are specified in terms of the radial frequency. The use of such models is usually restricted to image signals that are also cylinder-symmetric. Channelised Hotelling observers have been used with good success to predict the performance of human observers in detection tests.

As an example, Fig. 18 shows the CHO performance (detectability index (d′)) with dense of difference of Gaussian for an 8 mm sphere at 20 HU of the QRM 401 phantom and three $\mathrm{CTDI_{vol}}$ levels acquired in GE HD 750 CT scanner.

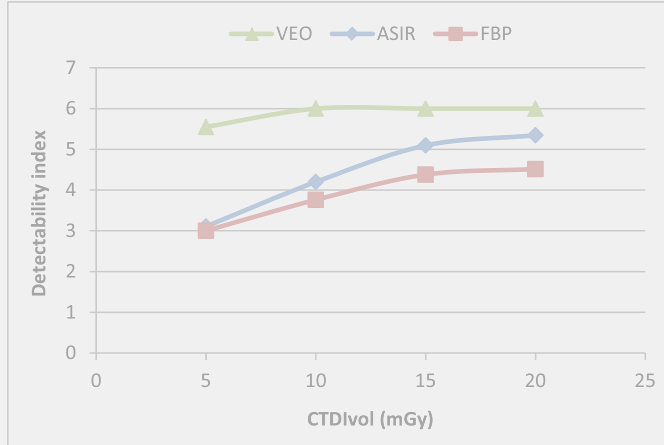

Figure 18. Detectability index (d_A) as a function of $\mathrm{CTDI_{vol}}$ for the different algorithms for a sphere of Ø 8 mm and a contrast with background of 20 HU in the QRM 401 Abdomen Phantom (QRM, Moehrendorf, Germany).

Agreement between observers

The first step to compare model observers and model/human observers is to have the same metrics to measure their performance. For a specific task, background, signal and model the investigator must choose between the area under the curve (AUC), sensitivity/specificity pairs, the percent correct (PC), the signal to noise ratio (SNR) or the detectability index (d′), then a comparison is possible.

Kappa test

To measure the agreement between observers it is common to use the Kappa coefficient. When observers are two or more the inter-observer variation can be computed. The Kappa test is based on the difference between the observer agreement (percentage where observers agree among themselves) and the expected agreement (agreement obtained just by chance). The formula for the Kappa test is then as follows:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{50}$$

where p_0 is the relative observed agreement among reviewers, and p_e is the probability of chance agreement.

The Kappa scale ranges from −1 to 1. 1 represents a perfect agreement, 0, the agreement is obtained just by chance, and −1 represents systematic disagreement. A generic scale proposed by Landis and Koch is used to help the investigator to interpret the Kappa coefficient (Table 3) [112].

The Kappa coefficient estimated itself could be obtained just by chance, so a P value can be calculated to interpret the result of the Kappa test. The P value is sensitive to sample size, so another Kappa test can be used to interpret the result, the weighted Kappa assigns weighting more or less important to different categories, to focus on categories where the difference is significant. But the weighting is defined by the investigator and the expert can disagree on the tuning of the weighted Kappa. The Kappa test is used to interpret the agreement, but this test is affected by the prevalence of the disease [50] (Fig. 19); in many cases a low Kappa test does not reflect a low agreement. Moreover, the Kappa test can give strange results when the observers have a high degree of agreement and when they are close to PC = 1.

Table 3
Genetic scale investigator to interpret the Kappa coefficient.

0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

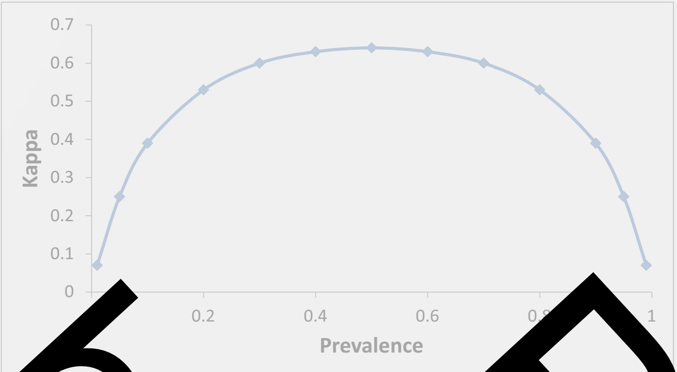

Figure 19. Kappa coefficient in terms of prevalence.

Bland–Altman plot

A Bland–Altman plot is often used to compare results between model observers and human observers [113]. When both observers measure the same parameter (i.e. d′ or PC) with the same images most of the time the correlation is good [57,108]. A good correlation for two observers that measure the same parameter does not imply a good agreement between the two observers.

A Bland–Altman plot shows the mean of the two observers in the abscissa, and the difference between the two observers in the ordinate. The limits of agreement are defined by the mean of the difference and the standard deviation of the difference. If a method is the gold standard then d represents the bias whereas if any methods are standard, d represents only systematic differences. Figure 20 shows an example comparing the performance of the NPWE model and human observers for a given detection task.

Conclusion and perspectives

Since the introduction of CT many efforts have been made to balance image quality with patient exposure. Image quality was first assessed using signal detection theory, and basic parameters such as image noise and spatial resolution, which made it possible to evaluate the strengths and weaknesses of acquisition protocols. With the technological developments of CT it became necessary to assess units in order to objectively enhance the benefit of new technological solutions. Global figures of merit of image quality were derived, still using signal theory functions, normalising the result by a standardised dose indicator: the $CTDI_{vol}$. If this approach seems enticing one has to remember that the use of one number to judge image quality is a simplified solution that can lead to false conclusions. Moreover, image quality assessment methods based on signal theory only do not include a clinically relevant task. With this kind of approach one could optimise aiming at getting the best theoretical image quality, rather than ensuring that images convey the relevant clinical information to make a correct diagnosis. In such a context, image quality assessment in the field of medical imaging should be task oriented and clinically relevant.

The use of mathematical model observers may be an appropriate solution, opening a way forward, even if the tasks investigated remain very simple and far from clinical reality. As shown in the review, there are several types of model observers, and the choice of a single solution might not be optimal. The disadvantage of model observers is that they are defined for simple situations, like the detection of a representative signal in a given phantom, and surely do not cover the whole range of characteristics that define image quality at the clinical level. This drawback can nonetheless become an advantage because their calculation can be kept relatively simple; they are objective and compatible with new image reconstruction techniques such as iterative reconstruction. They also lead to reproducible results which can be representative of human perception but avoid the burden of actual studies with human observers. They could be used to compare clinical protocols in terms of image quality and dose levels to initiate an optimisation process. Nevertheless, more studies should be performed in the future on correlation between model observer outcomes and human diagnostic accuracy.

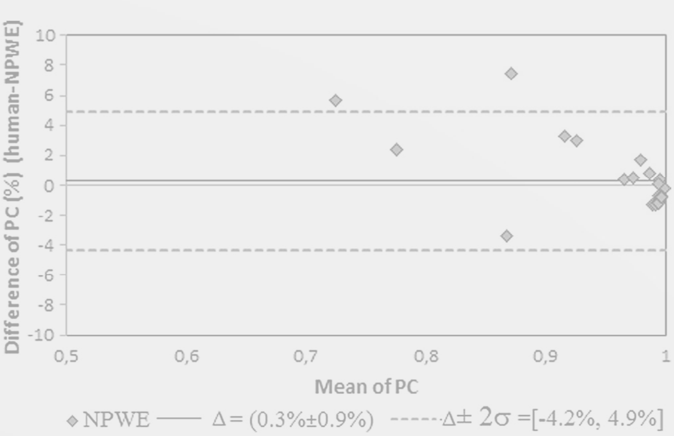

Figure 20. Bland–Altman plot of proportion correct (PC) difference between human and NPWE for 1% contrast and all mAs. The straight line represents the average difference (Δ) and the dash lines, the range of the differences [Δ ± 2σ], where σ is the standard deviation of the differences. The NPWE model was corrected by an efficiency of 0.38.

Acknowledgements

This work was supported by the German Radiation Protection Agency (BfS – UFO-Plan [illegible] 3[illegible]S[illegible]0007) and the Swiss National Science Foundation for research (FNS 320030_140995/1). The authors are thankful to R. van Engen from LRCB Nijmegen and Pr H. Bosmans from Katholieke Universiteit Leuven, and P. Monnin from CHUV, Lausanne, for their help for improving the structure of the manuscript.

References

[1] Thornton J. NCRP report no. 160: ionizing radiation exposure of the population of the United States. Phys Med Biol 2010;55:6327. doi:10.1088/0031-9155/55/20/6327.
[2] Aubert B, Sinno-Tellier S, Etard C. Exposition de la population française aux rayonnements ionisants liée aux actes de diagnostic médical en 2007. Institut de Radioprotection et de Sûreté Nucléaire et the Institut de Veille Sanitaire:
[3] Samara ET, Aroua A, Bochud FO, Ott B, Theiler T, Treier R, et al. Exposure of the Swiss population by medical x-rays: 2008 review. Health Phys 2012;102:263–70.
[4] Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety. Bernhard-Ströl C, Hachenberger C, Trugenberger-Schnabel A, Loebke-Reinl A, Peter J, editors. Umweltradioaktivitaet und strahlenbelastung. Annual report. 2012.
[5] IRSN. Exposition de la population française aux rayonnements ionisants liés aux actes de diagnostic médical en 2012. Rapport PRP-HOM N°2014-6. 2012. <http://www.irsn.fr/FR/Actualites_presse/Communiques_et_dossiers_de_presse/Pages/20141013_Rapport-Expri-Exposition-rayonnements-diagnostic-medical.aspx#.VOHGuuK1Z8E> [accessed 16.02.15].
[6] Boone J, Strauss K, Cody D, McCollough C, McNitt-Gray M, Toth T. Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations. Report of AAPM Task Group 204. 2011.
[7] Nguyen TTA, Le HND, Vo M, Wang Z, Luu L, Ramella-Roman JC. Three-dimensional phantoms for curvature correction in spatial frequency domain imaging. Biomed Opt Express 2012;3:1200–14. doi:10.1364/BOE.3.001200.
[8] Mironov V, Boland T, Trusk T, Forgacs G, Markwald RR. Organ printing: computer-aided jet-based 3D tissue engineering. Trends Biotechnol 2003;21:157–61. doi:10.1016/S0167-7799(03)00033-7.

[9] Solomon J, Samei E. Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. Med Phys 2014;41:091908. doi:10.1118/1.4893497.

[10] Tapiovaara MJ, Wagner R. SNR and DQE analysis of broad spectrum X-ray imaging. Phys Med Biol 1985;30:519. doi:10.1088/0031-9155/30/6/002.

[11] Nickoloff EL, Riley R. A simplified approach for modulation transfer function determinations in computed tomography. Med Phys 1985;12:437–42.

[12] Boone JM. Determination of the presampled MTF in computed tomography. Med Phys 2001;28:356–60.

[13] Judy PF. The line spread function and modulation transfer function of a computed tomographic scanner. Med Phys 1976;3:233–6.

[14] Nakaya Y, Kawata Y, Niki N, Umetatni K, Ohmatsu H, Moriyama N. A method for determining the modulation transfer function from thick microwire profiles measured with x-ray microcomputed tomography. Med Phys 2012;39[illegible]347–64. doi:1[illegible]18/1.4729711.

[15] Thorn[illegible] MM, Flynn MJ. Measurement of the spatial resolution [illegible] clin[illegible] vol[illegible]ric computed tomography scanner using a sphere pha[illegible], vol. 6142 [illegible]. p. [illegible]21Z doi:10.1117/12.654969.

[1[illegible]] [illegible]rimmer R, [illegible]ause J, Karolczak M, Lapp R, Kachelriess M. [illegible]essm[illegible] of spatial resolution i[illegible]CT. IEEE Nucl Sci Symp Conf Rec 200[illegible]62–6. do[illegible]0.1109/NSSMIC.20[illegible]4774508. 2008 NSS 08.

[1[illegible]] Friedma[illegible] N, Fu[illegible] GSK, Siewerdsen JH, Tsui BMW. A simple approach to [illegible] comput[illegible] tomography (CT) modulation transfer function (MTF) and noise-power sp[illegible]rum (NPS) using the American College of Radiology (ACR) accreditation [illegible]antom. Med Phys 2013;40:051907. doi:10.1118/1.48007[illegible].

[18] International [illegible]mmis[illegible] Radiation Units and Measurements. ICRU [illegible]ort No. 87: radi[illegible] and image-quality assessment in computed tom[illegible]aphy. J [illegible] 20[illegible]:1–149. doi:10.1093/jicru/ndt007.

[19] Mi[illegible] F, Beaumont S, Torfeh T, Gudinchet F, Verdun FR. Computed tomography commissioning programmes: how to obtain a reliable Mtf with an automatic approach? Radiat [illegible]etry 2010;ncq050. doi:10.1093/rpd/ncq050.

[20] Dainty JC, Shaw R. Image s[illegible]ce: princip[illegible], analysis and evaluation of photographic-type imaging [illegible]cesses. Acad[illegible]ic Press; 1974.

[21] Thibault J-B, Sauer KD, Boum[illegible] CA, Hsieh J. A three-dimensional statistical approach to improved image [illegible]for m[illegible]lical CT. Med Phys 2007;34:4526–44. doi:10.1118/1.2[illegible]499.

[22] Hsieh J, Nett B, Yu Z, Sauer K, Thibault J-[illegible]ouman CA. [illegible]cent advances in CT image reconstruction. Curr Radiol Rep [illegible]13;1:39–51. [illegible]10.1007/s40134-012-0003-7.

[23] Richard S, Li X, Yadava G, Samei E. Predictive [illegible]odels f[illegible]server [illegible]formance in CT: applications in protocol optimization[illegible]961. [illegible]p. [illegible]610H doi:10.1117/12.877069.

[24] Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Ver[illegible] FR. Update o[illegible] the non-prewhitening model observer in compute[illegible]mography f[illegible] the assessment of the adaptive statistical and model-b[illegible] ite[illegible]tive recon[illegible]ction algorithms. Phys Med Biol 2014;59:4047–64. d[illegible]10.108[illegible]/59/4/4047.

[25] Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Objective assessm[illegible] of image quality and dose reduction in CT iterative reconstruction. Med [illegible]ys 2014;41:071904. doi:10.1118/1.4881148.

[26] Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Tow[illegible]ds tas[illegible] assessment of CT performance: system and object MTF a[illegible] different reconstruction algorithms. Med Phys 2012;39:4115–22. doi:10.1118/1.4725171.

[27] Brunner CC, Abboud SF, Hoeschen C, Kyprianou IS. Signal detection and location-dependent noise in cone-beam computed tomography using the spatial definition of the Hotelling SNR. Med Phys 2012;39:3214–28. doi:10.1118/1.4718572.

[28] Miéville FA, Bolard G, Bulling S, Gudinchet F, Bochud FO, Verdun FR. Effects of computing parameters and measurement locations on the estimation of 3D NPS in non-stationary MDCT images. Phys Med 2013;29:684–94. doi:10.1016/j.ejmp.2012.07.001.

[29] Edyvean S. Understanding image quality and dose. ImPACT Feb. 2007. <http://www.impactscan.org/slides/course07/lect9/frame.htm> [accessed 29.04.15].

[30] Seeram E. Computed tomography: physical principles, clinical applications, and quality control, 3e. 2nd ed. Saunders; 2000.

[31] Edyvean S. The relationship between image noise and spatial resolution of CT scanners, <http://www.ctug.org.uk/meet02/noiseandspatialresct.pdf>; 2002 [accessed 11.03.15].

[32] Edyvean S, Keat N. Comparison of CT scanner image noise, image width, dose and spatial resolution using standard test methods. <http://www.aapm.org/meetings/04AM/pdf/14-2350-75226.pdf>; 2004 [accessed 11.03.15].

[33] Brooks RA, Di Chiro GD. Statistical limitations in x-ray reconstructive tomography. Med Phys 1976;3:237–40. doi:10.1118/1.594240.

[34] Riederer SJ, Pelc NJ, Chesler DA. The noise power spectrum in computed X-ray tomography. Phys Med Biol 1978;23:446–54.

[35] Atkinson JK. The quantitative assessment of CT scanners. University of London; 1980.

[36] Edyvean S. ImPACT MDA Report type testing of CT scanners – methods and MEthodology, <http://www.impactscan.org/reports/MDA9825.htm>; 1998 [accessed 11.03.15].

[37] Fuchs T, Kalender WA. On the correlation of pixel noise, spatial resolution and dose in computed tomography: theoretical prediction and verification by simulation and measurement. Phys Med 2003;XIX(2):153–64.

[38] Kalender WA. Computed tomography: fundamentals, system technology, image quality, applications. 3rd ed. Erlangen: Publicis; 2011.

[39] Bassano DA. Specification and quality assurance for CT scanners. AAPM Summer Sch 1980.

[40] Allisy-Roberts P, Williams JR. Farr's physics for medical imaging. Elsevier Health Sciences; 2007.

[41] Edyvean S. ImPACT NHS PASA 16 slice CT scanner comparison report version 14. <http://www.impactscan.org/reports/Report06012.htm>; 2006 [accessed 11.03.15].

[42] Edyvean MS. A methodical approach for comparison of CT scanner image quality relative to dose. Radiological Society of North America 2003 Scientific Assembly and Annual Meeting, November 30–December 5, 2003, Chicago, IL. <http://archive.rsna.org/2003/3107396.html> [accessed 29.04.15].

[43] Platten D, Keat N, Lewis M, Barret J, Edyvean S. ImPACT MHRA 04045 Toshiba 16 Report, page 26, <http://www.impactscan.org/reports/MHRA04045.htm>; 2003 [accessed 30.03.15].

[44] Nagel H-D. CT dose efficiency parameters. European medical ALARA network. WG 1: optimisation of patient exposure in CT procedures, synthesis document 2012. <https://www.yumpu.com/en/document/view/42382097/wg1-synthesis-report-pdf-a-11-mb-european-medical-alara-/5>; 2004 [accessed 17.02.15].

[45] Chao EH, Toth TL, Williams EC, Fox SH, Carleton CA, Bromberg NB. A statistical method of defining low contrast detectability, poster presented at RSNA Meeting; 2000.

[4[illegible]] Rose A. Vision: human and electronic. New York: Plenum Press; 1973.

[47] Torgersen GR, Hol C, Møystad A, Hellén-Halme K, Nilsson M. A phantom for [illegible]lified image quality control of dental cone beam computed tomography unit[illegible] Oral Surg Oral Med Oral Pathol Oral Radiol 2014;118:603–11. doi:1[illegible]016/j.oooo.2014.08.003.

[48] Nage[illegible]-D. Methoden zur bestimmung der dosiseffizienz von CT-scannern, pre[illegible]ed a[illegible]eeting; 2008.

[49] M[illegible]son LC[illegible]thods for the evaluation of image quality: a review. Radiat [illegible]ot Dosi[illegible]y [illegible]00;90:89–99.

[50] Beutel [illegible]undel HL, Van Metter RL. Handbook of medical imaging: physics and [illegible]chophysic[illegible] Press; 2000.

[51] Sa[illegible] E, Bad[illegible], Cha[illegible]borty D, Compton K, Cornelius C, Corrigan K, et al. Ass[illegible]men[illegible]play pe[illegible]rmance for medical imaging systems: executive summary [illegible]AAP[illegible]TG18 [illegible]port. Med Phys 2005;32:1205–25. doi:10.1118/1.1861159.

[52] Likert R. [illegible]echnique[illegible] the measurement of attitudes. Arch Psychol 1932;22(14[illegible].

[53] Jamieson S. Likert scales: h[illegible]o (ab)u[illegible] them. Med Educ 2004;38:1217–18. doi:10.1111/j.1365-2929[illegible]04.02012.x.

[54] Norman G. Likert scales[illegible]evels of meas[illegible]ement [illegible]he "laws" of statistics. Adv Health Sci Educ 20[illegible]15:625–32. do[illegible]10[illegible]s10459-010-9222-y.

[55] International Commissio[illegible] Radiation U[illegible] Measurements. Receiver operating characteristic ana[illegible] medic[illegible]aging, vol. 79. ICRU Rep. N°79. Bethesda (MD): International Co[illegible]ssion on Radiation Units and Measurements; 2008.

[5[illegible]] Samei E, Krupinski E. The handb[illegible]k of m[illegible]al imag[illegible]perception and te[illegible]hniques. Cambridge UK: Cambridg[illegible] Uni[illegible]ess; 20[illegible].

[57] [illegible]rrett HH, Myers KJ. Foundations of ima[illegible]scienc[illegible]Hob[illegible]en (NJ): Wiley-Inter[illegible]nce; 2[illegible]4.

[5[illegible]] Vennart W. ICRU Report 54: medical imag[illegible] – the a[illegible]essme[illegible] image quality [illegible]ography 1996;3:243–4. doi:10.10[illegible]8-817[illegible]9.

[59] Way T[illegible]an H-P, Hadjiiski L, Sahiner B, Chughtai A, Song T[illegible] al. Comp[illegible]er-ai[illegible] diagnosis of lung nodules on CT scans: ROC st[illegible] of its effec[illegible] [illegible]ologi[illegible]mance. Acad Radiol 2010;17:32[illegible]32. doi:10.[illegible]/j.acra.2[illegible].10.016.

[60] Li F, [illegible]yama M, Sh[illegible]chi J, Abe H, Li G, Suzuki K, et [illegible]logists' perfo[illegible]ance for di[illegible]ntiatin[illegible]ign from malignant lun[illegible]odules on [illegible]igh-[illegible]olution C[illegible]ng com[illegible]er-estimated likelihood of malignancy. Am [illegible]Roent[illegible] 20[illegible];183:120[illegible].

[61] [illegible]umura M[illegible] T, Kai[illegible], [illegible]re JW, McNitt-Gray M, Katada K. Effect of e[illegible]ing ada[illegible] image [illegible]er on low-contrast detectability in CT systems: applicatio[illegible] ROC analysis[illegible] J Biomed Imaging 2008;2008:379486. doi:10.1155/200[illegible]79486.

[62] Bushberg JT, Seibert JA, Leidh[illegible] EM, B[illegible] JM. The essential physics of medical imaging. Philadelph[illegible](PA): Wol[illegible]uwer Health; 2011.

[63] Wunderlich A, Abbey [illegible]Utility [illegible] rationale for choosing observer performance assessment paradig[illegible] detection tasks in medical imaging. Med Phys 2013;40:111903. doi:[illegible]118/1.4823755.

[64] Chakraborty DP. A brief history [illegible]FROC paradigm data analysis. Acad Radiol 2013;20:915–19. doi:10.1016/j.ac[illegible].2013.03.001.

[65] Popescu LM. Nonparametric signal detectability evaluation using an exponential transformation of the FROC curve. Med Phys 2011[illegible]702. doi:10.1118/1.3633938.

[66] Burgess A. Comparison of receiver operating characteris[illegible] and fo[illegible]choice observer performance-measurement methods. Med P[illegible]ys 19[illegible];22:64[illegible]55. doi:10.1118/1.597576.

[67] Macmillan NA, Creelman CD. Detection theory: a user[illegible]guid[illegible] Mahwah ([illegible]): Lawrence Erlbaum Associates; 2005.

[68] Bochud FO, Abbey CK, Eckstein MP. Visual signal de[illegible]ion [illegible]ctured backgrounds. III. Calculation of figures of merit for m[illegible]l observers [illegible] statistically nonstationary backgrounds. J Opt Soc Am A Op[illegible] Vis 2000;17:193–205. doi:10.1364/JOSAA.17.000193.

[69] Swets JA. Signal detection and recognition by human observers: contemporary readings. New York: Wiley; 1964.
[70] Craven BJ. A table of d′ for M-alternative odd-man-out forced-choice procedures. Percept Psychophys 1992;51:379–85.
[71] Hacker MJ, Ratcliff R. A revised table of d′ for M-alternative forced choice. Percept Psychophys 1979;26:168–70. doi:10.3758/BF03208311.
[72] Dahlquist G, Björck Å. Numerical methods. New York: Courier Corporation; 2012.
[73] Green DM, Swets JA. Signal detection theory and psychophysics. New York: John Wiley and Sons; 1966.
[74] Burgess AE. Visual perception studies and observer models in medical imaging. Semin Nucl Med 2011;41:419–36. doi:10.1053/j.semnuclmed.2011.06.005.
[75] Tapiovaara M. Efficiency of low-contrast detail detectability in fluoroscopic imaging. Med Phys 1997;24:655–64. doi:10.1118/1.598076.
[76] Eckst[illegible] MP, Abbey CK, Bochud FO. Visual signal detection [illegible]tured b[illegible]grounds. IV. Figures of merit for model performan[illegible] m[illegible]ple-[illegible]rnative forced-choice detection tasks with cor[illegible]ed respo[illegible]s. J Opt [illegible]oc Am A Opt Image Sci Vis 2000;17:2[illegible] doi:10.13[illegible] JOSAA.1[illegible]000206.
[77] Zhang Y, [illegible]eng S, Yu L, Carter RE, McCollough CH. [illegible]elation b[illegible]en hu[illegible] and m[illegible] observer performance for discrimination task in CT. P[illegible]ol 201[illegible]3389[illegible]04. doi:10.1088/0031-9155/59/13/3389.
[78] [illegible]andez-[illegible]on I, Geleijns J, Calzado A, Veldkamp WJH. Automate[illegible] assessment [illegible]low contrast sensitivity for CT systems using a model obse[illegible] Med Phys [illegible]11;38:S25[illegible]5. doi:10.1118/1.3577757.
[79] Yu L, Len[illegible], Chen [illegible]er JM, Carter RE, McCollough CH. Prediction [illegible]uman observe[illegible]ance in a 2-alternative forced choice low-contra[illegible]detection task[illegible] channelized Hotelling observer: impact of radiation dose [illegible] reconstruction algorithms. Med Phys 2013;40:041908. doi:10.1118/ 1.4794498.
[80] Monnin P, Marshall NW, Bos[illegible]chud FO, Verdun FR. Image quali[illegible] assessment in digital mam[illegible]graphy: pa[illegible]I. NPWE as a validated alternative for contrast detail analy[illegible] Phys Med B[illegible] 2011;56:4221–38. doi:10.1088/ 0031-9155/56/14/003.
[81] Borasi G, Samei E, Berto[illegible] M, Nitrosi A, Tassoni D. Contrast-detail analysis of three flat panel detecto[illegible] radi[illegible]Med Phys 2006;33:3580. doi:10.1118/1.2337636.
[82] Rivetti S, Lanconelli N, Bertolini M, Acc[illegible]ppati D. A ne[illegible]linical unit for digital radiography based on a thick amo[illegible]hous seleniu[illegible]plate: physical and psychophysical characterization. Me[illegible]Phys 2011;[illegible]4480–8. doi:10.1118/ 1.3605471.
[83] Wagner RF, Brown DG, Pastel MS. Applica[illegible] infor[illegible]y to the assessment of computed tomography. Med Phys 197[illegible]83–94.
[84] Myers KJ, Rolland JP, Barrett HH, Wagner RF. A[illegible]ure optimiz[illegible]on for emission imaging: effect of a spatially varying [illegible]round. J Opt[illegible] Am A 1990;7:1279–93.
[85] Burgess AE, Ghandeharian H. Visual sign[illegible] detecti[illegible]al-loca[illegible] identification. J Opt Soc Am A 1984;1:906–10. doi:10.1364/JOSAA.1.00[illegible]6.
[86] Burgess A. Visual signal detection. III. On Bayesian use of prior knowle[illegible] and cross correlation. J Opt Soc Am A 1985;2:1498–507.
[87] Wagner RF, Myers KJ, Tapiovaara MJ, Brown DG, Burgess AE. Sch[illegible] editor. Maximum a-posteriori detection and figures of [illegible]erit f[illegible]ection under uncertainty. 1990. p. 195–204 doi:10.1117/12.1879[illegible]
[88] Chesters MS. Human visual perception and ROC methodology in medical imaging. Phys Med Biol 1992;37:1433. doi:10.1088/0031-9155/37/7/001.
[89] Brown DG, Insana MF, Tapiovaara M. Detection performance of the idea[illegible] decision function and its McLaurin expansion: signal position unknown. J Acoust Soc Am 1995;97:379–98.
[90] Bochud FO, Abbey CK, Bartroff J, Vodopich D, Eckstein MP. Effect of the number of locations in MAFC experiments performed with mammograms 1999.
[91] Burgess AE. Evaluation of detection model performance in power-law noise, vol. 4324. 2001. p. 123–32 doi:10.1117/12.431180.
[92] Kotre CJ. The effect of background structure on the detection of low contrast objects in mammography. Br J Radiol 1998;71:1162–7. doi:10.1259/ bjr.71.851.10434911.
[93] Bochud FO, Valley JF, Verdun FR, Hessler C, Schnyder P. Estimation of the noisy component of anatomical backgrounds. Med Phys 1999;26:1365–70.
[94] Marshall NW, Kotre CJ, Robson KJ, Lecomber AR. Receptor dose in digital fluorography: a comparison between theory and practice. Phys Med Biol 2001;46:1283–96.
[95] Tseng H-W, Fan J, Kupinski MA, Sainath P, Hsieh J. Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers. Med Phys 2014;41:071910. doi:10.1118/1.4881143.
[96] Samei E, Richard S. Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology. Med Phys 2015;42:314–23. doi:10.1118/1.4903899.
[97] Chen B, Ramirez Giraldo JC, Solomon J, Samei E. Evaluating iterative reconstruction performance in computed tomography. Med Phys 2014;41:121913. doi:10.1118/1.4901670.
[98] Tapiovaara MJ, Wagner RF. SNR and noise measurements for medical imaging: I. A practical approach based on statistical decision theory. Phys Med Biol 1993;38:71. doi:10.1088/0031-9155/38/1/006.
[99] Rolland JP, Barrett HH. Effect of random background inhomogeneity on observer detection performance. J Opt Soc Am A 1992;9:649–58. doi:10.1364/ JOSAA.9.000649.
[100] Tapiovaara MJ. SNR and noise measurements for medical imaging. II. Application to fluoroscopic X-ray equipment. Phys Med Biol 1993;38:1761. doi:10.1088/0031-9155/38/12/006.
[101] Boedeker KL, Cooper VN, McNitt-Gray MF. Application of the noise power [illegible]pectrum in modern diagnostic MDCT: part I. Measurement of noise power [illegible]ctra and noise equivalent quanta. Phys Med Biol 2007;52:4027–46. d[illegible]10.1088/0031-9155/52/14/002.
[102] B[illegible]gess AE. Statistically defined backgrounds: performance of a modified [illegible]prew[illegible]ng observer model. J Opt Soc Am A Opt Image Sci Vis [illegible]994;11[illegible]7–42.
[103] Loo L[illegible] Metz CE. A comparison of physical image quality indices and ob[illegible]er performance in the radiographic detection of nylon beads. Phys Med [illegible] 1984;29:[illegible]56.
[104] [illegible]ang Y, [illegible] BT, [illegible]kstein MP. Evaluation of internal noise methods for [illegible]telli[illegible]erver [illegible]dels. Med Phys 2007;34:3312–22. doi:10.1118/ 1.275[illegible]3.
[105] Brank[illegible]JG. Eva[illegible] channelized Hotelling observer with internal-noise model[illegible] train-te[illegible]aradigm for cardiac SPECT defect detection. Phys Med Biol 201[illegible]7159–82. d[illegible]/0031-9155/58/20/7159.
[106] Leng S, Yu L, Zhang Y, [illegible]ter R, [illegible]dano AY, McCollough CH. Correlation between model obs[illegible]er and hum[illegible] observer performance in CT imaging when lesion locatio[illegible]s uncertain. [illegible]d Phys [illegible]3;40:081908. doi:10.1118/ 1.4812430.
[107] Brankov JG. Optimiza[illegible] of the internal [illegible]odels for channelized Hotelling observer. 2011 IEEE Int[illegible] Biome[illegible]aging Nano Macro 2011;1788–91. doi:10.1109/ISBI.2011.5872753.
[108] Hernandez-Giron I, Calzado A, [illegible]eleijns [illegible]MS, Veldkamp WJH. Comparison between human an[illegible]odel o[illegible]er perfo[illegible]ance in low-contrast detection tasks in CT images: appl[illegible]ation [illegible]s reco[illegible]ructed with filtered back projection and iterative algorit[illegible]s. Br J [illegible]diol [illegible]014;87:20140014. [illegible]10.12[illegible]jr.20140014.
[109] Myers KJ, Barrett HH. Addition of a chan[illegible] mechani[illegible] to the [illegible]al-observer mod[illegible] Opt Soc Am A 1987;4:2447–57. do[illegible]64/JO[illegible]7.
[110] A[illegible]y CK, Barrett HH. Human- and model-observer [illegible]ormance [illegible]ra[illegible] [illegible]ectrum noise: effects of regularization and object v[illegible]ability. J Opt [illegible]m A Opt [illegible]Vis 2001;18:473–88.
[111] Ba[illegible] HH, Yao [illegible]Rolland JP, Myers KJ. Model obse[illegible]rs for as[illegible]ment of i[illegible]e quality. Pr[illegible]atl Acad Sci U S A 1993;90:9758–[illegible]
[112] L[illegible]dis JR, Koch [illegible]he mea[illegible]ment of observer agreeme[illegible]or categorical d[illegible] Biometri[illegible]977;33:[illegible]–74.
[113] Bla[illegible]M, [illegible]an DG. [illegible]eement between methods of measurement with multip[illegible]servati[illegible]per[illegible]dividual. J Biopharm Stat 2007;17:571–82. [illegible]80/105[illegible]0701329[illegible]2.

10 Discussion and perspectives

In recent years, the work of medical physicists has been devoted to controlling the compliance of radiological units using dose indicators as surrogates for patient exposure and physical measurements as surrogates for image quality. With this approach, radiological units were classified according to their efficacy when it came to converting X-rays into information based on engineering or physical criteria. With linear systems, some link could be established between the physical metrics and the clinical images. However, limiting image quality assessments to the first level of the Fryback scale is in sufficient and it can even be counterproductive when dealing with non linear systems, because the linear relationship between dose and image quality properties, as well as the spatial invariance of the statistical properties of the signal, is removed. In that case, nice looking images can be obtained in a wide range of doses in clinical routines; but a nice image does not necessarily mean that the diagnostic information is preserved. As a result, there is now strong pressure to reduce the dose used in CT. In parallel, manufacturers are promoting new data processing which enables massive dose reductions whilst claiming to preserve image quality. All of this means that it is imperative to control the impact of dose reduction on image quality. This PhD thesis developed and evaluated new methodologies to assess the image quality of modern CT systems with newer metrics because traditional physical methods used to assess image quality assessment are known to be insufficient to properly assess IR. The aim of these new methodologies is to establish a link with radiologists to offer a more clinically relevant analysis (second level of the Fryback scale) of the impact of IR on image quality. In this context we proposed to place image quality as the main parameter in clinical practice optimisation followed by a control of the required patient exposure.

The work done in the framework of this PhD demonstrated that medical physicists need additional skills if they want to be part of quality assessment but also involved in the patient dose optimisation process. Firstly, the use of ideal model observers is useful to benchmark CT units or check the compliance of units. Secondly, anthropomorphic model observers can be used to benchmark and optimise clinical CT protocols because they establish a link to human performance. During this PhD thesis, we used these two kinds of MO to benchmark CT units and clinical protocols and this methodology with model observers is now applicable within a clinical context, but some aspects of our approach must still be refined in order to go beyond certain limitations. The advantage of model observers that work in the image domain is that it is possible to directly compare model observer performance and human performance. A major disadvantage, however, in a clinical context is that their outcome is robust only with a certain number of acquisitions. Furthermore, CT exams provide volumetric data and radiologists have to perform many tasks in their daily practice to make a diagnosis. For the time being, it is impossible to define metrics that could encompass all aspects of image quality, but in the framework of this study we focussed on low-contrast detection (in a 2D mode) that is of essential for certain abdominal and brain CT indications in the context of patient dose reduction. Using this simple task (detection task of a simple signal in a homogenous background), human performance is similar between single and multi-slice assessments [90]. In the future we propose to define some clinical indications where radiologists would translate the clinical image quality requirements into a set of simple task-based image quality criteria. Then, medical physicists would ensure that the task-based image criteria are reached when the chosen protocol is used in realistic situations like in a structured background or with moving structures for example.

The majority of this PhD thesis is concentrated on low contrast detectability, but we tried to generalise the methodology for high-contrast parameters, even if using a detection task with a high-contrast lesion is not the best way to follow because the detection of a high contrast lesion is not a major issue, unless the dose is very low and the noise component is at a high frequency (in that case the signal can be confused with noise). In high contrast paradigm one challenge is to move towards a characterization task when dealing with high contrast structures. This paradigm could be evaluated using an estimation task like shape discrimination or size estimation in the image domain (e.g. when a variation of 20% in the size of a lesion appears on an image during patient follow up; it is vital to be sure that this variation is real.) One of the challenges of benchmarking CT units is defining equivalent image acquisition and reconstruction parameters. A consensus has been reached in the framework of this project to work in a comparable framework of $CTDI_{vol}$ values but CTDI and DLP do not benefit from a primary metrology standard, and are therefore subject to large error margins. For example, under Swiss law, the tolerated error margin for the measurement of CTDI is of ±20 %. This implies that two examinations performed on two separate units, although physically delivering the same dose to two patients, may display doses showing up to 40 % difference. With the implementation of ATCM, the dose is managed completely differently from one manufacturer to another. This constitutes a limitation regarding the possibility of actual inter-manufacturer comparisons. At the same time, the CTDI tolerance could be reduced with the technological development and the optimization of clinical protocols and patient dose monitoring could be better estimated on the basis of the values displayed by the CT units. A metrological attachment of the model observer measurement method will provide the medical physicist a task-based approach to establish image quality requirements depending on clinical questions defined in collaboration with radiologists. As standards exist for dose measurement and physical metrics (like NPS and MTF), it is important to develop a standard for model observers.

In conclusion, the role of medical physicists in radiology is changing and much time has to be devoted to assessing image quality properties that matter for patient care as well as fully understanding the needs of radiologists. This thesis focuses on the relationship between dose and image quality but the image quality is largely impacted by a broader clinical context (i.e. temporal properties of the acquisitions as well as the timing condition of the contrast media injection). This means that medical physicists have an important role in the terms of continued education to ensure that the technological aspects that improve image quality are correctly understood and employed by all users in a clinical context. Diagnostic accuracy can be maximized by optimizing technical aspects of CT acquisitions but also by making an optimal use of the CT machines in their clinical environment (Figure 22).

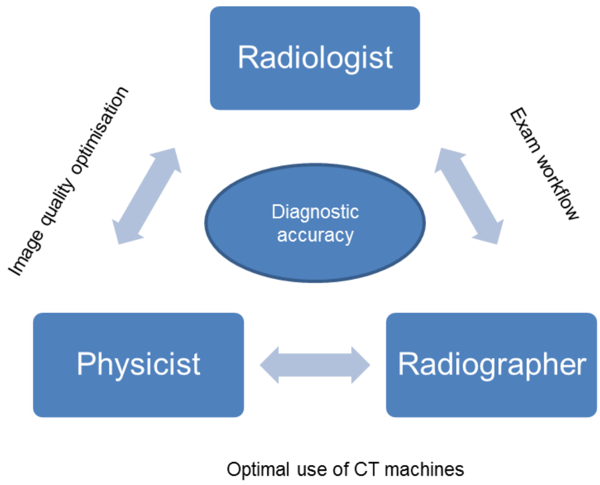

Figure 22: Diagnostic accuracy a multi-disciplinary inter dependency

References

[1] Cindy, *NCRP Report 160| Bethesda, MD*, NCRP Bethesda MD (n.d.).

[2] B. Aubert, S. Sinno-Tellier, and C. Etard, Exposition de la population française aux rayonnements ionisants liée aux actes de diagnostic médical en 2007, in Institut de Radioprotection et de Sûreté Nucléaire et the Institut de Veille Sanitaire(2010).

[3] E.T. Samara, A. Aroua, F.O. Bochud, *et al.*, Exposure of the Swiss population by medical x-rays: 2008 review, Health Phys. **102**(3), 263–270 (2012).

[4] C. Ströl, C. Hachenberger, A. Trugenberger-Schnabel, and J. Peter, *Umweltradioaktivität und Strahlenbelastung*, (2010).

[5] R.L. Coultre, J. Bize, M. Champendal, *et al.*, Exposure of the Swiss Population by Radiodiagnostics: 2013 Review, Radiat. Prot. Dosimetry ncv462 (2015).

[6] IRSN, *Exposition de la population française aux rayonnements ionisants liés aux actes de diagnostic médical en 2012. Rapport PRP-HOM N°2014-6.*, (2012).

[7] A.-F. Perez, C. Devic, C. Colin, and N. Foray, [The low doses of radiation: Towards a new reading of the risk assessment], Bull. Cancer (Paris) **102**(6), 527–538 (2015).

[8] J.D. Mathews, A.V. Forsythe, Z. Brady, *et al.*, Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians, BMJ **346**, f2360 (2013).

[9] H. Hricak, D.J. Brenner, S.J. Adelstein, *et al.*, Managing radiation use in medical imaging: a multifaceted challenge, Radiology **258**(3), 889–905 (2011).

[10] D.J. Durand, A rational approach to the clinical use of cumulative effective dose estimates, AJR Am. J. Roentgenol. **197**(1), 160–162 (2011).

[11] D.J. Brenner, I. Shuryak, and A.J. Einstein, Impact of reduced patient life expectancy on potential cancer risks from radiologic imaging, Radiology **261**(1), 193–198 (2011).

[12] D.J. Brenner, R. Doll, D.T. Goodhead, *et al.*, Cancer risks attributable to low doses of ionizing radiation: assessing what we really know, Proc. Natl. Acad. Sci. U. S. A. **100**(24), 13761–13766 (2003).

[13] K. Rothkamm, S. Balroop, J. Shekhdar, P. Fernie, and V. Goh, Leukocyte DNA damage after multi-detector row CT: a quantitative biomarker of low-level radiation exposure, Radiology **242**(1), 244–251 (2007).

[14] L. Walsh, R. Shore, A. Auvinen, T. Jung, and R. Wakeford, Risks from CT scans--what do recent studies tell us?, J. Radiol. Prot. Off. J. Soc. Radiol. Prot. **34**(1), E1-5 (2014).

[15] *ICRP: ICRP Publication 103*, (n.d.).

[16] D.G. Fryback and J.R. Thornbury, The efficacy of diagnostic imaging, Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak. **11**(2), 88–94 (1991).

[17] H.H. Barrett and K.J. Myers, *Foundations of image science* (Wiley-Interscience, 2004).

[18] R.F. Wagner, C.E. Metz, and G. Campbell, Assessment of medical imaging systems and computer aids: a tutorial review, Acad. Radiol. **14**(6), 723–748 (2007).

[19] *J.B. Solomon, "Advanced techniques for image quality assessment of modern x-ray computed tomography systems." PhD Dissertation, Duke University Graduate School, March 2016.*, (n.d.).

[20] W.A. Kalender, *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications* (John Wiley & Sons, 2011).

[21] *International Electrotechnical Committee 1994. Medical diagnostic X-ray equipment - Radiation conditions for use in the determination of characteristics.*, Stand. IEC 61267 (2015).

[22] R.L. Dixon, A new look at CT dose measurement: beyond CTDI, Med. Phys. **30**(6), 1272–1280 (2003).

[23] E.J. Hall and D.J. Brenner, Cancer risks from diagnostic radiology: the impact of new epidemiological data, Br. J. Radiol. **85**(1020), e1316-1317 (2012).

[24] W. Huda and F.A. Mettler, Volume CT dose index and dose-length product displayed during CT: what good are they?, Radiology **258**(1), 236–242 (2011).

[25] C. Borrás, W. Huda, and C.G. Orton, Point/counterpoint. The use of effective dose for medical procedures is inappropriate, Med. Phys. **37**(7), 3497–3500 (2010).

[26] *impactscan.org | ctdosimetry.xls - ImPACT's ct dosimetry tool*, (n.d.).

[27] J.-P.J. Yu, A.P. Kansagra, D.M. Naeger, R.G. Gould, and F.V. Coakley, Template-driven computed tomography radiation dose reporting: implementation of a radiology housestaff quality improvement project, Acad. Radiol. **20**(6), 769–772 (2013).

[28] *J. Boone, K. Strauss, D. Cody, C. McCollough, M. McNitt-Gray, and T. Toth, "Size-specific dose estimates (SSDE) in pediatric and adult body CT exams," Report of AAPM Task Group 204, 2011.*,.

[29] M. van Straten, P. Deak, P.C. Shrimpton, and W.A. Kalender, The effect of angular and longitudinal tube current modulations on the estimation of organ and effective doses in x-ray computed tomography, Med. Phys. **36**(11), 4881–4889 (2009).

[30] H. Greess, A. Nömayr, H. Wolf, *et al.*, Dose reduction in CT examination of children by an attenuation-based on-line modulation of tube current (CARE Dose), Eur. Radiol. **12**(6), 1571–1576 (2002).

[31] W.A. Kalender, H. Wolf, and C. Suess, Dose reduction in CT by anatomically adapted tube current modulation. II. Phantom measurements, Med. Phys. **26**(11), 2248–2253 (1999).

[32] H. Greess, H. Wolf, U. Baum, *et al.*, Dose reduction in computed tomography by attenuation-based on-line modulation of tube current: evaluation of six anatomical regions, Eur. Radiol. **10**(2), 391–394 (2000).

[33] S.T. Schindera, R.C. Nelson, T. Yoshizumi, *et al.*, Effect of automatic tube current modulation on radiation dose and image quality for low tube voltage multidetector row CT angiography: phantom study, Acad. Radiol. **16**(8), 997–1002 (2009).

[34] S. Rizzo, M. Kalra, B. Schmidt, *et al.*, Comparison of angular and combined automatic tube current modulation techniques with constant tube current CT of the abdomen and pelvis, AJR Am. J. Roentgenol. **186**(3), 673–679 (2006).

[35] R.K. Kaza, J.F. Platt, M.M. Goodsitt, *et al.*, Emerging techniques for dose optimization in abdominal CT, Radiogr. Rev. Publ. Radiol. Soc. N. Am. Inc **34**(1), 4–17 (2014).

[36] M.A. Habibzadeh, M.R. Ay, A.R.K. Asl, H. Ghadiri, and H. Zaidi, Impact of miscentering on patient dose and image noise in x-ray CT imaging: phantom and clinical studies, Phys. Medica PM Int. J. Devoted Appl. Phys. Med. Biol. Off. J. Ital. Assoc. Biomed. Phys. AIFB **28**(3), 191–199 (2012).

[37] D. Gandhi, D.J. Crotty, G.M. Stevens, and T.G. Schmidt, Technical Note: Phantom study to evaluate the dose and image quality effects of a computed tomography organ-based tube current modulation technique, Med. Phys. **42**(11), 6572–6578 (2015).

[38] L. Yu, H. Li, J.G. Fletcher, and C.H. McCollough, Automatic selection of tube potential for radiation dose reduction in CT: a general strategy, Med. Phys. **37**(1), 234–243 (2010).

[39] Y.J. Suh, Y.J. Kim, S.R. Hong, *et al.*, Combined use of automatic tube potential selection with tube current modulation and iterative reconstruction technique in coronary CT angiography, Radiology **269**(3), 722–729 (2013).

[40] J.-B. Thibault, K.D. Sauer, C.A. Bouman, and J. Hsieh, A three-dimensional statistical approach to improved image quality for multislice helical CT, Med. Phys. **34**(11), 4526–4544 (2007).

[41] A.K. Hara, R.G. Paden, A.C. Silva, J.L. Kujak, H.J. Lawder, and W. Pavlicek, Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study, AJR Am. J. Roentgenol. **193**(3), 764–771 (2009).

[42] A.C. Silva, H.J. Lawder, A. Hara, J. Kujak, and W. Pavlicek, Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm, AJR Am. J. Roentgenol. **194**(1), 191–199 (2010).

[43] D. Marin, R.C. Nelson, S.T. Schindera, *et al.*, Low-tube-voltage, high-tube-current multidetector abdominal CT: improved image quality and decreased radiation dose with adaptive statistical iterative reconstruction algorithm--initial clinical experience, Radiology **254**(1), 145–153 (2010).

[44] Y. Sagara, A.K. Hara, W. Pavlicek, A.C. Silva, R.G. Paden, and Q. Wu, Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients, AJR Am. J. Roentgenol. **195**(3), 713–719 (2010).

[45] K.T. Flicek, A.K. Hara, A.C. Silva, Q. Wu, M.B. Peter, and C.D. Johnson, Reducing the radiation dose for CT colonography using adaptive statistical iterative reconstruction: A pilot study, AJR Am. J. Roentgenol. **195**(1), 126–131 (2010).

[46] F. Pontana, J. Pagniez, T. Flohr, *et al.*, Chest computed tomography using iterative reconstruction vs filtered back projection (Part 1): Evaluation of image noise reduction in 32 patients, Eur. Radiol. **21**(3), 627–635 (2011).

[47] F. Pontana, A. Duhamel, J. Pagniez, *et al.*, Chest computed tomography using iterative reconstruction vs filtered back projection (Part 2): image quality of low-dose CT examinations in 80 patients, Eur. Radiol. **21**(3), 636–643 (2011).

[48] C. Ghetti, O. Ortenzia, and G. Serreli, CT iterative reconstruction in image space: a phantom study, Phys. Medica PM Int. J. Devoted Appl. Phys. Med. Biol. Off. J. Ital. Assoc. Biomed. Phys. AIFB **28**(2), 161–165 (2012).

[49] D. Fleischmann and F.E. Boas, Computed tomography--old ideas and new technology, Eur. Radiol. **21**(3), 510–517 (2011).

[50] Y. Funama, K. Taguchi, D. Utsunomiya, *et al.*, Combination of a low-tube-voltage technique with hybrid iterative reconstruction (iDose) algorithm at coronary computed tomographic angiography, J. Comput. Assist. Tomogr. **35**(4), 480–485 (2011).

[51] M. Kim, J.M. Lee, J.H. Yoon, *et al.*, Adaptive Iterative Dose Reduction Algorithm in CT: Effect on Image Quality Compared with Filtered Back Projection in Body Phantoms of Different Sizes, Korean J. Radiol. **15**(2), 195–204 (2014).

[52] F.A. Miéville, F. Gudinchet, F. Brunelle, F.O. Bochud, and F.R. Verdun, Iterative reconstruction methods in two different MDCT scanners: Physical metrics and 4-alternative forced-choice detectability experiments – A phantom approach, Phys. Med. **29**(1), 99–110 (2013).

[53] P.J. Pickhardt, M.G. Lubner, D.H. Kim, *et al.*, Abdominal CT With Model-Based Iterative Reconstruction (MBIR): Initial Results of a Prospective Trial Comparing Ultralow-Dose With Standard-Dose Imaging, Am. J. Roentgenol. **199**(6), 1266–1274 (2012).

[54] A. Neroladaki, D. Botsikas, S. Boudabbous, C.D. Becker, and X. Montet, Computed tomography of the chest with model-based iterative reconstruction using a radiation exposure similar to chest X-ray examination: preliminary observations, Eur. Radiol. **23**(2), 360–366 (2013).

[55] Verdun F.R. and Racine D., J.G. Ott, M.J. Tapiovaara, *et al.*, Image quality in CT: From physical measurements to model observers, Phys. Med. **31**(8), 823–843 (2015).

56 S. Richard, D.B. Husarik, G. Yadava, S.N. Murphy, and E. Samei, Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms, Med. Phys. **39**(7), 4115–4122 (2012).

57 J.M. Wilson, O.I. Christianson, S. Richard, and E. Samei, A methodology for image quality evaluation of advanced CT systems, Med. Phys. **40**(3), 031908 (2013).

58 E. Samei and E. Krupinski, *The Handbook of Medical Image Perception and Techniques* (Cambridge University Press, 2014).

59 T. Way, H.-P. Chan, L. Hadjiiski, *et al.*, Computer-Aided Diagnosis of Lung Nodules on CT Scans: ROC Study of Its Effect on Radiologists' Performance, Acad. Radiol. **17**(3), 323–332 (2010).

60 T. Way, H.-P. Chan, L. Hadjiiski, *et al.*, Computer-Aided Diagnosis of Lung Nodules on CT Scans: ROC Study of Its Effect on Radiologists' Performance, Acad. Radiol. **17**(3), 323–332 (2010).

61 A. Wunderlich and C.K. Abbey, Utility as a rationale for choosing observer performance assessment paradigms for detection tasks in medical imaging, Med. Phys. **40**(11), 111903 (2013).

62 W. Vennart, ICRU Report 54: Medical imaging—the assessment of image quality, Radiography **3**(3), 243–244 (1996).

63 A. BURGESS, Comparison of receiver operating characteristic and forced-choice observer performance-measurement methods, Med. Phys. **22**(5), 643–655 (1995).

64 N.A. Macmillan and C.D. Creelman, *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, 2005).

65 A. BURGESS, COMPARISON OF RECEIVER OPERATING CHARACTERISTIC AND FORCED-CHOICE OBSERVER PERFORMANCE-MEASUREMENT METHODS, Med. Phys. **22**(5), 643–655 (1995).

66 H.H. Barrett and K.J. Myers, *Foundations of image science* (Wiley-Interscience, 2004).

67 I. Hernandez-Giron, J. Geleijns, A. Calzado, and W.J.H. Veldkamp, Automated assessment of low contrast sensitivity for CT systems using a model observer, Med. Phys. **38**(S1), S25–S35 (2011).

68 L. Yu, S. Leng, L. Chen, J.M. Kofler, R.E. Carter, and C.H. McCollough, 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. **40**(4), 041908 (2013).

69 Y. Zhang, S. Leng, L. Yu, R.E. Carter, and C.H. McCollough, Correlation between human and model observer performance for discrimination task in CT, Phys. Med. Biol. **59**(13), 3389–3404 (2014).

70 J.G. Ott, F. Becce, P. Monnin, S. Schmidt, F.O. Bochud, and F.R. Verdun, 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. **59**(15), 4047 (2014).

71 K.J. Myers and H.H. Barrett, Addition of a channel mechanism to the ideal-observer model, J. Opt. Soc. Am. A **4**(12), 2447–2457 (1987).

72 H.H. Barrett, J. Yao, J.P. Rolland, and K.J. Myers, Model observers for assessment of image quality, Proc. Natl. Acad. Sci. **90**(21), 9758–9765 (1993).

73 J.Y. Vaishnav, W.C. Jung, L.M. Popescu, R. Zeng, and K.J. Myers, Objective assessment of image quality and dose reduction in CT iterative reconstruction, Med. Phys. **41**(7), 071904 (2014).

74 H.H. Barrett, C.K. Abbey, B.D. Gallas, and M.P. Eckstein, Stabilized estimates of Hotelling-observer detection performance in patient-structured noise, in (International Society for Optics and Photonics, 1998), pp. 27–44.

75 B.D. Gallas and H.H. Barrett, Validating the use of channels to estimate the ideal linear observer, J. Opt. Soc. Am. A **20**(9), 1725–1738 (2003).

[76] R. Zeng, N. Petrick, M.A. Gavrielides, and K.J. Myers, Approximations of noise covariance in multi-slice helical CT scans: impact on lung nodule size estimation, Phys. Med. Biol. **56**(19), 6223–6242 (2011).

[77] C.K. Abbey and H.H. Barrett, Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability, J. Opt. Soc. Am. A **18**(3), 473–488 (2001).

[78] M.P. Eckstein, C.K. Abbey, and J.S. Whiting, Human vs model observers in anatomic backgrounds, in (1998), pp. 16–26.

[79] A.E. Burgess and B. Colborne, Visual signal detection. IV. Observer inconsistency, J. Opt. Soc. Am. A **5**(4), 617–627 (1988).

[80] Z.L. Lu and B.A. Dosher, Characterizing human perceptual inefficiencies with equivalent internal noise, J. Opt. Soc. Am. A Opt. Image Sci. Vis. **16**(3), 764–778 (1999).

[81] J.G. Brankov, Evaluation of Channelized Hotelling Observer with Internal-Noise Model in a Train-Test Paradigm for Cardiac SPECT defect detection, Phys. Med. Biol. **58**(20), 7159–7182 (2013).

[82] S. Leng, L. Yu, Y. Zhang, R. Carter, A.Y. Toledano, and C.H. McCollough, Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain, Med. Phys. **40**(8), 081908 (2013).

[83] D. Racine, A.H. Ba, J.G. Ott, F.O. Bochud, and F.R. Verdun, Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer, Phys. Medica Eur. J. Med. Phys. **32**(1), 76–83 (2016).

[84] D. Racine, J.G. Ott, A. Ba, N. Ryckx, F.O. Bochud, and F.R. Verdun, Objective task-based assessment of low-contrast detectability in iterative reconstruction, Radiat. Prot. Dosimetry **169**(1–4), 73–77 (2016).

[85] D. Racine, A. Viry, F. Becce, *et al.*, Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners, Med. Phys. **44**(9), e153–e163 (2017).

[86] D. Racine, N. Ryckx, A. Ba, J.G. Ott, F.O. Bochud, and F.R. Verdun, Benchmarking of CT for patient exposure optimisation, Radiat. Prot. Dosimetry **169**(1–4), 78–83 (2016).

[87] D. Racine, N. Ryckx, A. Ba, *et al.*, *Towards a standardization of image quality in abdominal CT: Results from a multicentre study*, Submitt. Eur. Radiol. (n.d.).

[88] D. Rotzinger, D. Racine, K. Alfudhili, *et al.*, *TASK-BASED MODEL OBSERVER ASSESSMENT OF ASIR-V INFLUENCE ON OBJECTIVE IMAGE QUALITY IN ONCOLOGIC THORACIC MULTIDETECTOR CT*, Submitt. Investig. Radiol. (n.d.).

[89] D. Racine, J.G. Ott, P. Monnin, *et al.*, *Task-based assessment of impact of multiplanar reformations on objective image quality in iterative reconstruction in computed tomography*, Submitt. Radiol. (n.d.).

[90] A. Ba, D. Racine, A. Viry, F.R. Verdun, S. Schmidt, and F.O. Bochud, Low contrast detection in abdominal CT: comparing single-slice and multi-slice tasks, in (2017), p. 101360S–101360S–10.

Annexe 1: Development of the Signal to Noise Ratio of Prewhitening matched filter model observer

The general definition of the SNR is: $SNR^2 = \frac{[\lambda(g_1)-\lambda(g_2)]^2}{\sigma_\lambda^2}$

The general definition of the standard deviationσ_λ^2:

$\sigma^2 = \langle \lambda(g) - \langle \lambda(g)^2 \rangle \rangle$

$\sigma^2 = \langle \lambda^2(g) + \langle \lambda(g) \rangle^2 - 2\lambda(g)\langle \lambda(g) \rangle \rangle$

$\sigma^2 = \langle \lambda^2(g) \rangle + \langle \langle \lambda(g) \rangle^2 \rangle - \langle 2\lambda(g)\langle \lambda(g) \rangle \rangle$

$\sigma^2 = \langle \lambda^2(g) \rangle + \langle \lambda(g) \rangle^2 \langle \quad \rangle - 2\langle \lambda(g)\langle \lambda(g) \rangle \rangle$

$\sigma^2 = \langle \lambda^2(g) \rangle + \langle \lambda(g) \rangle^2 - 2\langle \lambda(g) \rangle^2$

$\sigma^2 = \langle \lambda^2(g) \rangle - \langle \lambda(g) \rangle^2$

The λ of PW model observer is equal to $\lambda(g) = s^T K_n^{-1} g$

The$\sigma_\lambda^2$of the PW is:

$\sigma^2 = \langle (s^T K_n^{-1} \bar{g})^2 \rangle - \langle (s^T K_n^{-1} \bar{g}) \rangle^2$

$\sigma^2 = [\langle s^T K_n^{-1} \bar{g} \rangle \langle \bar{g}^T K_n^{-1} s \rangle] - \langle s^T K_n^{-1} \bar{g} \rangle \langle \bar{g}^T K_n^{-1} s \rangle$

$\sigma^2 = \langle s^T K_n^{-1} \bar{g} \bar{g}^T K_n^{-1} s \rangle - s^T K_n^{-1} \langle \bar{g} \rangle \langle \bar{g}^T \rangle K_n^{-1} s$

$\sigma^2 = s^T K_n^{-1} \langle \bar{g} \bar{g}^T \rangle - \langle \bar{g} \rangle \langle \bar{g}^T \rangle K_n^{-1} s$

$\sigma^2 = s^T K_n^{-1} \langle \bar{g} \bar{g}^T \rangle - \langle \bar{g} \rangle \langle \bar{g}^T \rangle K_n^{-1} s$

$\sigma^2 = s^T K_n^{-1} (\langle \bar{g} \bar{g}^T \rangle - \langle \bar{g} \rangle \langle \bar{g}^T \rangle) K_n^{-1} s$

$\sigma^2 = s^T K_n^{-1} (K_n) K_n^{-1} s$

$\sigma^2 = s^T K_n^{-1} s$

The expression of the SNR of the PW model observer is development below

$$SNR^2 = \frac{\left[\overline{\lambda(g_1)} - \overline{\lambda(g_2)}\right]^2}{\sigma_\lambda^2}$$

$$SNR^2 = \frac{[s^T K_n^{-1} \overline{g_1} - s^T K_n^{-1} \overline{g_2}]^2}{\sigma_\lambda^2}$$

$$SNR^2 = \frac{[s^T K_n^{-1} \overline{g_1} - s^T K_n^{-1} \overline{g_2}]^2}{s^T K_n^{-1} s}$$

$$SNR^2 = \frac{[s^T K_n^{-1} (\overline{g_1} - \overline{g_2})]^2}{s^T K_n^{-1} s}$$

$$SNR^2 = \frac{[s^T K_n^{-1} s]^2}{s^T K_n^{-1} s}$$

$$SNR^2 = s^T K_n^{-1} s$$

Annexe 2: Development of the Signal to Noise Ratio of Non Prewhitening matched filter model observer

The λ of PM model observer is equal to $\lambda(g) = s^T g$

The$\sigma_{\lambda}^2$of the NPW is:

$$\sigma^2 = \langle (s^T \bar{g})^2 \rangle - \langle (s^T \bar{g}) \rangle^2$$
$$\sigma^2 = \langle s^T \bar{g} \bar{g}^T s \rangle - \langle s^T \bar{g} \rangle \langle \bar{g}^T s \rangle$$
$$\sigma^2 = s^T \langle \bar{g} \bar{g}^T \rangle s - s^T \langle \bar{g} \rangle \langle \bar{g}^T \rangle s$$
$$\sigma^2 = s^T (\langle \bar{g} \bar{g}^T \rangle - \langle \bar{g} \rangle \langle \bar{g}^T \rangle) s$$
$$\sigma^2 = s^T (\langle \bar{g} \bar{g}^T \rangle - \langle \bar{g} \rangle \langle \bar{g}^T \rangle) s$$
$$\sigma^2 = s^T (K_n) s$$

The expression of the SNR of the NPW model observer is development below:

$$SNR^2 = \frac{\left[\overline{\lambda(g_1)} - \overline{\lambda(g_2)}\right]^2}{\sigma_{\lambda}^2}$$
$$SNR^2 = \frac{[s^T \overline{g_1} - s^T \overline{g_2}]^2}{\sigma_{\lambda}^2}$$
$$SNR^2 = \frac{[s^T \overline{g_1} - s^T \overline{g_2}]^2}{s^T K_n s}$$

$$SNR^2 = \frac{[s^T \Delta g]^2}{s^T K_n s}$$

List of attended conferences

Only posters and oral communications as first author are indicated

Comparison of abdominal CT protocols: a multi-center study on image quality and radiation dose levels
Racine Damien, Ryckx Nick, Ba Alexandre, Viry Anaïs, Becce Fabio, Schmidt Sabine, Verdun Francis R.
56^{ièmes} Journées Scientifiques de la Société Française de Physique Médicale
Lyon, France, 15-16 juin 2017
(communication orale)

Evaluation of low contrast detectability and patient exposure using abdominal CT protocols: A multicentre study
Racine Damien, Ryckx Nick, Ba Alexandre, Viry Anaïs, Becce Fabio, Verdun Francis R., Schmidt Sabine
Journées scientifiques de la Société Suisse de Radiologie, 104ème congrès annuel
Bern, Suisse, 8-10 juin, 2017
(communication orale)

Optimization of abdominal CT protocols using a mathematical model observer
Racine Damien, Ryckx Nick, Viry Anaïs, Becce Fabio, Francis R. Verdun, Schmidt Sabine
European Congress of Radiology 2017
Vienne, Autriche, 1 - 5 mars 2017
(communication orale)

Comparaison objective de la résolution spatiale et de la détection à bas contraste en utilisant différents protocoles cliniques
Racine Damien, Viry Anaïs, Becce Fabio, Schmidt Sabine, Ba Alexandre, Bochud François O., Verdun Francis R.
33èmes journées des des Laboratoires Associés de Radiophysique et de Dosimétrie (L.A.R.D.)
Besançon, France, 13 - 14 Octobre 2016
(communication orale)

Benchmarking of abdominal CT protocols using a Channelized Hotelling Observer
Racine Damien, Ryckx Nick, Viry Anaïs, Becce Fabio, Schmidt Sabine, Francis R. Verdun
50th Annual Meeting of Swiss Society of Radiobiology and Medical Physics
Sursee, Suisse, 25 - 27 aout 2016
(poster)

Characterisation of CT units using a Dose Efficiency Index concept
Racine Damien, Monnin Pascal, Viry Anaïs, Ba Alexandre, Bochud François O., Schegerer Alexander, Edyvean Sue, Verdun Francis R.
4th International Conference on Image Formation in X-Ray Computed Tomography
Bamberg, Germany, 18 -22 juillet 2016
(poster)

Task-based evaluation of image quality with a wide-volume CT scanner: comparison of helical and axial acquisition modes
Racine Damien, Ryckx Nick, Ott Julien, Becce Fabio, Rotzinger David, Verdun Francis R.
Journées scientifiques de la Société Suisse de Radiologie, 103ème congrès annuel, congrès annuel
Davos, Suisse, 19-21 mai, 2016
(poster)

Impact of large X-ray beam collimation on image quality
Racine Damien, Ba Alexandre, Ott Julien, Bochud François O., Verdun Francis R.
SPIE Medical Imaging 2016 - Image perception, observer performance and technology assessment
San Diego, USA, 27 février – 3 mars, 2016
(communication orale)

Benchmarking de scanners médicaux pour optimiser l'exposition des patients
Racine Damien, Ba Alexandre, Ott Julien, Ryckx Nick, Bochud François O., Verdun Francis R.
32èmes journées des Laboratoires Associés de Radiophysique et de Dosimétrie (L.A.R.D.)
Gradignan, France, 09-10 novembre 2015
(communication orale)

Towards an objective way to assess image quality in CT
Racine Damien, Ryckx Nick, Ba Alexandre, Ott Julien, Bochud François O., Verdun Francis R.
Journées scientifiques de la Société Suisse de Radiologie, 102ème congrès annuel, congrès annuel
Bâle, Suisse, 4-6 juin, 2015
(poster)

Objective task based assessment of low contrast detectability in iterative Reconstruction
Racine Damien, Ba Alexandre, Ott Julien, Ryckx Nick, Bochud François O., Verdun Francis R.
Optimisation in X-ray and Molecular Imaging 2015 (OXMI 2015)
Gothenburg, Suède, 28-30 mai 2015
(communication orale)

Benchmarking of CT to patient exposure optimization
Racine Damien, Ba Alexandre, Ott Julien, Ryckx Nick, Bochud François O., Verdun Francis R.
Optimisation in X-ray and Molecular Imaging 2015 (OXMI 2015)
Gothenburg, Suède, 28-30 mai 2015
(communication orale)

Evaluation objective de la détection à bas contraste en scanner
Racine Damien, Ba Alexandre, Ott Julien, Bochud François O., Verdun Francis R.
31èmes journées des Laboratoires Associés de Radiophysique et de Dosimétrie (L.A.R.D.)
CEA Saclay, France, 13-14 Octobre 2014
(communication orale)

List of publications

Peer-reviewed, first author

<u>Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners</u>
Racine Damien, Viry Anaïs, Becce Fabio, Ba Alexandre, Bochud François O., Edyvean Sue, Schegerer Alexander, Verdun Francis R.
Med Phys. 2017 Sep;44(9):e153-e163. doi: 10.1002/mp.12224.

<u>BENCHMARKING OF CT FOR PATIENT EXPOSURE OPTIMISATION</u>
Racine Damien, Ryckx Nick, Ba Alexandre, Ott Julien, Bochud François O., Verdun Francis R.
Radiat Prot Dosimetry March 2, 2016 doi:10.1093/rpd/ncw021

<u>OBJECTIVE TASK-BASED ASSESSMENT OF LOW-CONTRAST DETECTABILITY IN ITERATIVE RECONSTRUCTION</u>
Racine Damien, Ba Alexandre, Ott Julien, Bochud François O., Verdun Francis R.
Radiat Prot Dosimetry February 27, 2016 doi:10.1093/rpd/ncw020.

<u>Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer</u>
Racine Damien, Ba Alexandre, Ott Julien, Bochud François O., Verdun Francis R.
Phys Med. 2016 Jan;32(1):76-83. doi: 10.1016/j.ejmp.2015.09.011.

<u>Image quality in CT: From physical measurements to model observers</u>
Verdun Francis R.* and Racine Damien*, Ott Julien G., Tapiovaara Markku J., Toroi Paula, Bochud François O., Veldkamp Wouter J., Schegerer Alexander, Bouwman Ramona W., Giron Irene H., Marshall Nicholas W., Edyvean Sue
Phys Med. 2015 Dec; 31(8):823-43. doi: 10.1016/j.ejmp.2015.08.007.
*Co-auteurs

Peer-reviewed, co-author

<u>Assessment of low contrast detection in CT using model observers: Developing a clinically-relevant tool for characterising adaptive statistical and model-based iterative reconstruction</u>
Ott Julien G, Ba Alexandre, Racine Damien, Viry Anaïs, Bochud François O., Verdun Francis R.
Z Med Phys. 2017 Jun; 27(2):86-97. doi: 10.1016/j.zemedi.2016.04.002. Epub 2016 May 4.

<u>Dual-Energy CT: Basic Principles, Technical Approaches, and Applications in Musculoskeletal Imaging (Part 1).</u>
Omoumi Patrick, Becce Fabio, Racine Damien, Ott Julien G., Andreisek Gustav, Verdun Francis R.
Semin Musculoskelet Radiol. 2015 Dec;19(5):431-7. doi: 10.1055/s-0035-1569253. Epub 2015 Dec 22.

<u>Optimization of Radiation Dose and Image Quality in Musculoskeletal CT: Emphasis on Iterative Reconstruction Techniques (Part 1)</u>
Omoumi Patrick, Becce Fabio, Ott Julien G., Racine Damien, Verdun FR.
Semin Musculoskelet Radiol. 2015 Dec;19(5):415-21. doi: 10.1055/s-0035-1569255. Epub 2015 Dec 22.

Patient exposure optimisation through task-based assessment of a new iterative reconstruction technique: the ADMIRE algorithm
Ott Julien G., Ba Alexandre H., Racine Damien, Ryckx Nick, Bochud François O., Alkadhi Hatem, Verdun Francis R.
Radiat Prot Dosimetry. 2016 Mar 8. pii: ncw019. [Epub ahead of print]

Anthropomorphic model observer performance in three-dimensional detection task for low-contrast computed tomography
Ba Alexandre, Eckstein Miguel P., Racine Damien, Ott Julien G., Verdun Francis R., Kobbe-Schmidt Sabine, Bochud François O.
J Med Imaging (Bellingham).2016 Jan; 3(1):011009. doi: 10.1117/1.JMI.3.1.011009.Epub 2015 Dec 29

Conferences proceedings, first author

Characterization CT unit using a dose efficiency index concept
Racine Damien, Monnin Pascal, Bochud François O., Viry Anaïs, Schegerer Alexander, Edyvean Sue, Verdun Francis R.
Proceeding "The 4th International Conference on Image Formation in X-Ray Computed Tomography"

Impact of large X-ray beam collimation on image quality
Racine Damien, Ba Alexandre H., Ott Julien G., Bochud François O., Verdun Francis R.
Proc. SPIE 9787, Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment, 978714; doi:10.1117/12.2217018

Conferences proceedings, co-author

Characterization of a CT unit for the detection of low contrast structures
Viry, Anaïs, Racine Damien, Ba Alexandre, Becce Fabio, Bochud François O., Verdun, Francis R.
Proc. SPIE 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment, 101361C (2017/03/10); doi: 10.1117/12.2250529

Low contrast detection in abdominal CT: Comparing single-slice and multi-slice tasks
Ba Alexandre, Racine Damien, Viry Anaïs, Verdun Francis R., Schmidt Sabine, Bochud François O.
Proc. SPIE 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment, 101360S (2017/03/10); doi: 10.1117/12.2254237

Low contrast detectability in CT for human and model observer in multi-slice data sets
Ba Alexandre, Racine Damien, Ott Julien H., Eckstein Miguel P., Verdun Francis R., Bochud François O.
Proc. SPIE 9416, Medical Imaging 2015: Image Perception, Observer Performance, and Technology Assessment, 94160F (2015/03/17); doi: 10.1117/12.2082009

Under review, first author

Towards a standardization of image quality in abdominal CT: Results from a multicentre study
Racine Damien, Ryckx Nick, Ba Alexandre, Becce Fabio, Viry Anaïs, Verdun Francis R., Schmidt Sabine
Submitted in European Radiology

Task-based assessment of impact of multiplanar reformations on objective image quality in iterative reconstruction in computed tomography
Racine Damien, Ott Julien G., Monnin Pascal, Rotzinger David, Omoumi Patrick, Dugert Eric, Verdun Francis R., Becce Fabio
Submitted in Radiology

Under review, co-author

TASK-BASED MODEL OBSERVER ASSESSMENT OF ASIR-V INFLUENCE ON OBJECTIVE IMAGE QUALITY IN ONCOLOGIC THORACIC MULTIDETECTOR CT
Rotzinger David, Racine Damien, Alfudhili Khalid, Keller Nathalie, Verdun Francis R., Beigelman-Aubry Catherine, Becce Fabio
Submitted in Investigative Radiology

Curriculum Vitae

RACINE Damien

ETAT CIVIL

- 29 ans
- Nationalité Française
- Marié, 1 enfant
- Permis B
- Permis de séjour B

COORDONNEES

+41 (0)78 644 89 06
racine.damien@gmail.com
Adresse :
Avenue du 24 janvier, 2b
1020 RENENS
Suisse

LANGUES

- Français :
  langue maternelle
- Anglais :
  niveau intermédiaire

INFORMATIQUE

- Python
- Visual Basic
- Pack Office
- Photoshop,
- Lightroom
- ImageJ
- Scilab
- R
- Labview

CENTRES D'INTERETS

- Cyclisme
- Photographie (3 ans)

FORMATION

Actuellement — **Thèse : Caractérisation des systèmes de CT cliniques et optimisation des protocoles basés sur des observateurs mathématiques**
Institut de Radiophysique - Lausanne

Octobre 2014 — **Expert en radioprotection**

2012-2014 — **Master Ingénieries pour la Santé et le Médicament**
Parcours : Physique Médicale
Université Joseph Fourier - Grenoble

2011-2012 — **Licence 3 Physique Chimie**
Parcours Physique, option physique nucléaire Université Franche-Comté - Besançon

Juin 2011 — **Formation APAVE : Personne Compétente en Radioprotection.** Module théorique

2010-2011 — **Licence Professionnelle Dosimétrie et Radioprotections Médicales**
IUT Belfort-Montbéliard

2009-2010 — **DUT Mesures Physiques**
accès directement en 2ème année
IUT Belfort-Montbéliard

2008-2009 — **PCEM1 (1ère année de médecine)**
Université Franche-Comté - Besançon

EXPERIENCES PROFESSIONNELLES

2014 — Stage, **CHUV - IRA - Lausanne**
Février à Août
- Validation et automatisation de l'utilisation de modèles observateurs (NPWE, CHO) pour le contrôle qualité en CT, dans le cadre de l'article 74.

2013 — Stage, **Service de médecine nucléaire - C.H. Haguenau**
Avril à Juin
- Développement d'outils sur ImageJ pour l'exploitation d'images issues du contrôle qualité.

2012 Projet, **Laboratoire Chrono-Environnement - Besançon**

Janvier à Mai

- Test de référence du code Monte-Carlo MCNPX pour l'étude des particules secondaires suite à des irradiations aux ions lourds

2011 Dosimétriste**, Service de radiothérapie**
C.H.U. Jean Minjoz - Besançon
Juillet à Août

- Radiothérapie conformationelle 3D (Oncentra Master Plan)

2011 Stage, **Service de radiothérapie**
C.H.U. Jean Minjoz - Besançon
Mars à Juin

- Service radiothérapie : Etude de l'impact de l'utilisation de produits de contraste sur la dosimétrie, réalisation dosimétrie conformationnelle 3D, initiation aux contrôles qualité
- Service médecine nucléaire : Etude de la gestion des déchets radioactifs en milieu hospitalier

Projet, **C.H. Montbéliard**
Janvier à Mars

- Etude de la norme NFC 15-160 : Règles générales sur les installations pour la production et l'utilisation de rayons X