RESEARCH ARTICLE

Dose-efficiency quantification of computed tomography systems using a model-observer

Maximilian Göppel[1] | **Mathias Anton[2]** | **Hugo de las Heras Gala[1]** | **Augusto Giussani[1]** | **Sebastian Trinkl[1]** | **Bernhard Renger[3]** | **Gunnar Brix[1]**

[1]Department of Medical and Occupational Radiation Protection, Federal Office for Radiation Protection, Neuherberg, Germany

[2]Department of Dosimetry for Radiation Therapy and Diagnostic Radiology, Physikalisch-Technische Bundesanstalt, Berlin, Germany

[3]Department of Diagnostic and Interventional Radiology, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

Correspondence
Maximilian Göppel, Federal Office for Radiation Protection, Ingolstädter Landstraße 1, 85764 Oberschleißheim, Germany.
Email: mgoeppel@bfs.de

Abstract

Background: Recent advances in computed tomography (CT) technology have considerably improved the quality of CT images and reduced radiation exposure in patients. At present, however, there is no generally accepted figure of merit (FOM) for comparing the dose efficiencies of CT systems.

Purpose: (i) To establish an FOM that characterizes the quality of CT images in relation to the radiation dose by means of a mathematical model observer and (ii) to evaluate the new FOM on different CT systems and image reconstruction algorithms.

Methods: Images of a homogeneous phantom with four low-contrast inserts were acquired using three different CT systems at three dose levels and a representative protocol for CT imaging of low-contrast objects in the abdomen. The images were reconstructed using filtered-back projection and iterative algorithms. A channelized hotelling observer with difference-of-Gaussian channels was applied to compute the detectability (d'). This was done for each insert and each of the considered imaging conditions from square regions of interest (ROIs) that were (semi-)automatically centered on the inserts. The estimated detectabilities (d') were averaged in the first step over the three dose levels ($\langle d' \rangle$), and subsequently over the four contrast inserts ($\langle d' \rangle_w$). All calculation steps included a dedicated assessment of the related uncertainties following accepted metrological guidelines.

Results: The determined detectabilities (d') varied considerably with the contrast and diameter of the four inserts, as well as with the radiation doses and reconstruction algorithms used for image generation ($d' = 1.3$–5.5). Thus, the specification of a single detectability as an FOM is not well suited for comprehensively characterizing the dose efficiency of a CT system. A more comprehensive and robust characterization was provided by the averaged detectabilities $\langle d' \rangle$ and, in particular, $\langle d' \rangle_w$. Our analysis reveals that the model observer analysis is very sensitive to the exact position of the ROIs.

Conclusions: The presented automatable software approach yielded with the weighted detectability $\langle d' \rangle_w$ an objective FOM to benchmark different CT systems and reconstruction algorithms in a robust and reliable manner. An essential advantage of the proposed model-observer approach is that uncertainties in the FOM can be provided, which is an indispensable prerequisite for type testing.

Maximilian Göppel and Mathias Anton contributed equally to this work.

- -

1 | INTRODUCTION

Computed tomography (CT) is a common and widely used imaging technology for the diagnosis and early detection of diseases, as well as therapy planning and monitoring.[1] The annual number of CT examinations performed has steadily increased over the past few years. According to the latest UNSCEAR report,[1] the contribution of CT to the frequency of all X-ray examinations is relatively small (approximately 9.9 %) in high-income countries, while it provides the largest contribution to the collective effective dose from all X-ray examinations in these countries (approximately 60%).

Parallel to this development, several hardware and software innovations, such as tube-current modulation, low-voltage imaging, spectral shaping, dynamic collimation, photon-counting detectors, and iterative image reconstruction algorithms, have improved image quality and/or decreased the radiation dose to patients per examination.[2–5] This corresponds to the optimization principle of radiation protection according to which "doses should […] be kept as low as reasonably achievable, taking into account economic and societal factors.[6]"

The radiation dose related to a CT scan is characterized by the standardized CT dose index (*CTDI*) and the dose length product. However, there is no internationally accepted standard for characterizing the quality of CT images, even though there are some approaches, such as those by Riederer et al.[7] and Brooks and Di Chiro[8] used by the British CT scanner evaluation center (ImPACT). A key problem in characterizing the quality of CT images is that the imaging system is neither linear nor shift invariant. This is because the data pre-processing steps and image reconstruction operate in a non-linear way.[9] Image-quality metrics,[10,11] such as the signal-to-noise ratio, contrast-to-noise ratio, noise power spectrum, and modulation transfer function require system linearity and shift-invariance. Consequently, the suitability of these metrics is severely limited.[12–15]

Alternatively, the diagnostic image quality can be defined in terms of how well small structures can be detected in a certain setting.[16] This task-based assessment can be performed by either a human or a model observer. Using this approach, a figure of merit (FOM) that quantifies the image quality for a specific diagnostic task, imaging system, and observer can be defined. As human-observer studies, which typically entail the evaluation of "receiver operating characteristic" (ROC) curves, are very time consuming[17] and costly, mathematical model observers can be employed instead. The growing interest in the model-observer approach has been supported by a recent multicenter study[18] and an inter-laboratory comparison.[19] As a result, a new

IEC standard[20] permits model observers to be used for acceptance and constancy tests on CT scanners.

The aim of our investigation was to continue the aforementioned developments in model observers to establish an FOM that characterizes the quality of CT images in relation to the radiation dose, thus making it possible for use in type testing of CT systems. Type testing is usually performed on a few specimens of a device model at the manufacturer's site as part of the technical specifications of the model series. The approach presented in this paper is based on the detectability[12] of low-contrast objects within a homogeneous phantom and is tested at three different CT scanners. To meet the practical requirements of a type-testing procedure, the development was accompanied by a group of experts in the field, including representatives of the authorities in the field of medical radiation protection and manufacturers.

2 | THEORY

2.1 | Description of the task-based model observer

The quality of CT images can be quantified using task-based quality assessments, where tasks can refer to the reading and interpretation of suspicious structures by either a human or a mathematical observer. A rather simple task is the detection of a lesion, which can be formulated as a binary classification (present/absent) and further limited to the case of a "signal known exactly" and "background known exactly" situation, following the convention of Barrett et al.[16] This study deals with this setting only. A review of task-based measures for an image-quality assessment was conducted by Barrett et al.[16]

In general, a scalar observer-response variable z derived from an image vector f of length n, where n is the number of image pixels, is generated by

$$z = w(\boldsymbol{f}) + \epsilon, \tag{1}$$

where the scalar function $w(f)$ is the response to the image f, and ϵ is the internal noise contribution to the observer response.[21] This response is based on a mathematical model commonly referred to as a model observer.[16,22] The different proposed model observers vary in their choice of the functional form for $w(f)$ and a probability distribution for ϵ. Internal noise can be used to tune the performance of model observers to that of human observers.[21,23] In the present study, internal noise was not considered because the model observer was intended to perform optimally.[24,25]

Image f is classified as belonging to class 1 (signal-present) if z exceeds a chosen threshold value δ, and as belonging to class 0 (signal-absent) otherwise as follows:

$$z \geq \delta \rightarrow f \in \text{class 1}(z_1),$$
$$z < \delta \rightarrow f \in \text{class 0}(z_0). \tag{2}$$

Typically, a few hundred images were used, both with and without a test signal. If the resulting distributions of z_1 and z_0 are independent and normally distributed,[26,27] a detectability index d' can be defined by the first- and second-order statistics of the observer response variables:

$$d' = \frac{\mu_{z_1} - \mu_{z_0}}{\sqrt{0.5 \cdot \left(\sigma_{z_1}^2 + \sigma_{z_0}^2\right)}}. \tag{3}$$

In this equation, μ_{z_i} and $\sigma_{z_i}^2$ are the mean and variance of the corresponding distributions.[16,26] d' represents a measure of the separation between the centers of the distributions z_1 and z_0, relative to their average variance.

In the special case of a function (w) being linear in its argument (linear-model observer), Equation (1) can be written as:

$$z = w^\mathsf{T} f + \epsilon, \tag{4}$$

where w represents a vector of weights, commonly termed an observer template.

2.2 | The hotelling observer

The ideal linear classifier is the Hotelling observer (HO), which was introduced by Smith and Barrett[28] and Fiete et al.[29] and named after the statistician, Harold Hotelling.[30] Template w_{HO} of this observer depends on the mean image and covariance matrices of the two object classes considered (1 and 0). The classifier is ideal because it maximizes the detectability d' defined in Equation (3). We let μ_{f_1} be the mean signal-present image of all N_1 images of class 1, and μ_{f_0} be the mean signal-absent image of all N_0 images of class 0. The average covariance (V_f) can be defined using the covariances of both classes (V_{f_i}) as follows:

$$V_f = \frac{V_{f_1} + V_{f_0}}{2}. \tag{5}$$

The corresponding observer template is

$$w_{\text{HO}} = V_f^{-1} \left(\mu_{f_1} - \mu_{f_0}\right), \tag{6}$$

and the observer response is

$$z_{\text{HO}} = w_{\text{HO}}^\mathsf{T} f = \left(\mu_{f_1} - \mu_{f_0}\right)^\mathsf{T} V_f^{-1} f. \tag{7}$$

The multiplication of image f by V_f^{-1} "decorrelates" the noise in the image.[31] If the covariances for the two classes of images do not differ considerably (small-signal approximation), an estimate $\hat{V}_f$ of the covariance matrix can be used[32]:

$$\hat{V}_f = \frac{1}{N_0 + N_1 - 2} \left[\sum_{k=1}^{N_0} \left(\mu_{f_0} - f_{0,k}\right)\left(\mu_{f_0} - f_{0,k}\right)^\mathsf{T} + \sum_{l=1}^{N_1} \left(\mu_{f_1} - f_{1,l}\right)\left(\mu_{f_1} - f_{1,l}\right)^\mathsf{T} \right]. \tag{8}$$

2.3 | The channelized hotelling observer

The channelized hotelling observer (CHO) approximates the performance of an ideal linear observer (HO) and enables an objective image-quality assessment linked to the task of lesion detection. The details are described in the works by Abbey[33] and Abbey and Bochud.[34]

Myers and Barrett[35] introduced a channel mechanism for HO that reduces dimensionality and the subsequent computational burden. The channels are organized in a channel matrix U of size $n \times \nu$, where n is the number of image pixels, and ν is the number of channels. The image, covariance, and observer response can then be channelized as follows:

$$\tilde{f} = U^\mathsf{T} f, \tag{9}$$

$$\tilde{V}_f = U^\mathsf{T} V_f U, \tag{10}$$

$$z_{\text{CHO}} = \left(\mu_{\tilde{f}_1} - \mu_{\tilde{f}_0}\right)^\mathsf{T} \tilde{V}_f^{-1} \tilde{f}. \tag{11}$$

The tilde symbol ($\sim$) denotes the channelized versions of the image vectors and covariance matrix, which are reduced to length ν and size $\nu \times \nu$, respectively.

Following Wunderlich,[36] we used $\nu = 10$ channels in this study. Therefore, the size of the covariance matrix (V) was reduced to 10×10. Radially symmetric channels have been proposed to detect circular image structures (e.g., test structures in a phantom). In this study, we chose difference-of-Gaussian (DoG) channels:

$$U_j(x, y) = 2\pi\sigma_j^2 \left(Q^2 \exp\left(-2\pi^2 Q^2 \sigma_j^2 \cdot (x^2 + y^2)\right) - \exp\left(-2\pi^2 \sigma_j^2 \cdot (x^2 + y^2)\right) \right). \tag{12}$$

The x- and y-coordinates are given as multiples of the pixel size. The parameters σ_j ($\sigma_j = \sigma_0 \alpha^j$; $\sigma_0 = 0.005$; $\alpha = 1.4$) and $Q = 1.67$ were used as reported by Abbey and Barrett.[21] The overlapping (non-orthogonal)

channels were implemented with normalization using the Frobenius norm, which are presented in the toolbox by Wunderlich.[36]

2.4 | Calculation of an FOM characterizing the dose efficiency of CT systems

For meaningful characterization of CT-image quality, it is appropriate to determine the detectability of small, low-contrast structures in relation to the radiation dose. For this purpose, several rods with different contrast levels (c) and diameters (D) inserted into a phantom are frequently imaged at different dose levels (γ_i, $i = 1, \ldots, I$). Based on the acquired images, an FOM can be computed to describe the dose efficiency of the CT system. The starting point of the following considerations is the well-established detectability d' used in many publications.[9,32,37–39] We followed the approach used by Wunderlich[36,39] to estimate the performance of the infinitely-trained CHO. The concept presented in the present study is applicable to any number of rods and dose levels.

The d' values calculated for the different rods ($j = 1, \ldots, J$) imaged at different dose levels (γ_i, $i = 1, \ldots, I$) were averaged by approximating the integral over the considered γ_i values using the trapezoidal rule:

$$
\begin{aligned}
\langle d' \rangle_j &= \frac{\int_{\gamma_1}^{\gamma_I} d'(\gamma)\, d\gamma}{\int_{\gamma_1}^{\gamma_I} d\gamma} \\
&\approx \frac{1}{2(\gamma_I - \gamma_1)} \sum_{i=2}^{I} \left(d'_{i,j} + d'_{i-1,j} \right)(\gamma_i - \gamma_{i-1}) \\
&= \frac{1}{2(\gamma_I - \gamma_1)} \sum_{i=1}^{I} \kappa_i \cdot d'_{i,j},
\end{aligned}
\tag{13}
$$

where $\kappa_1 = \gamma_2 - \gamma_1, \ldots, \kappa_i = \gamma_{i+1} - \gamma_{i-1}(i = 2, \ldots, I-1)$, $\ldots, \kappa_I = \gamma_I - \gamma_{I-1}$ are the dose weights for the corresponding values of d'_i. In the final step, the weighted mean over all contrast rods $\langle d' \rangle_w$ is calculated as follows:

$$
\langle d' \rangle_w = \sum_{j=1}^{J} \lambda_j \langle d' \rangle_j .
\tag{14}
$$

The weights λ_j are given by the diameters (D_j) and contrasts (c_j) of the rods as follows:

$$
\lambda_j = \frac{(D_j \cdot c_j)^{-1}}{\sum_{k=1}^{J} (D_k \cdot c_k)^{-1}} .
\tag{15}
$$

This definition is based on the approximately linear dependency of the detectability d' on the diameter and contrast of the rods. Hernandez-Giron et al.[40] showed a linear increase of d' with object size; a simulation study (by M.A.) confirmed the linear dependence of d' on the contrast. The weights defined in Equation (15) are a sufficiently good approximation to the real situation to balance the effect of the different inserts and to avoid the dominance of the easily detectable inserts.

2.5 | Uncertainty analysis

Careful calculation of uncertainties is critical for defining proper acceptance conditions when a threshold is specified for a type-testing procedure. If not stated otherwise, expanded uncertainties giving a confidence level of approximately 95 % (coverage factor of 1.96) are reported in the present paper. The estimation of the uncertainty of $\langle d' \rangle_w$ requires an assessment of the uncertainties of its components λ and $\langle d' \rangle$, and thus also of the uncertainty of d'. For simplicity, the following presentation specifically refers to four contrast rods and three dose levels, which correspond to the measurements described below, following the order of calculation. Adaptations to other experimental setups are possible.

The uncertainties of the d' values are given by Wunderlich's exact confidence intervals for CHO.[39] However, these confidence intervals are not symmetric around the point estimate for d', making it difficult to perform subsequent uncertainty calculations according to the "Guide to the Expression of Uncertainty in Measurement" (GUM).[41] Therefore, we derived a symmetric standard uncertainty for d' based on the larger part of the confidence interval.

The uncertainty of $\langle d' \rangle$ obtained by following the GUM recommendations is as follows:

$$
\begin{aligned}
u^2 \left(\langle d' \rangle_j \right) =\ & \sum_{i=1}^{3} \left| \frac{\partial \langle d' \rangle_j}{\partial d'_{i,j}} \right|^2 \cdot u_i^2 \left(d'_{i,j} \right) \\
& + \sum_{i=1}^{3} \left| \frac{\partial \langle d' \rangle_j}{\partial \gamma_i} \right|^2 \cdot u_i^2 (\gamma_i) \\
& + 2 \cdot \sum_{i=1}^{3} \frac{\partial \langle d' \rangle_j}{\partial d'_{i,j}} \frac{\partial \langle d' \rangle_j}{\partial \gamma_i} \cdot \mathrm{cov} \left(\gamma_i, d'_{i,j} \right),
\end{aligned}
\tag{16}
$$

where the covariances $\mathrm{cov}(\gamma_i, d'_{k,j})$ for $i \neq k$ are neglected because the detectability in one acquisition does not depend on the dose applied in another.

The detectability d' is proportional to the dose γ and therefore, the correlation coefficient $\rho_{d',\gamma} \approx 1$. The covariances for $i = k$ can then be approximated by

$$cov\left(\gamma_i, d'_{i,j}\right) \approx \sqrt{var(\gamma_i)} \cdot \sqrt{var\left(d'_{i,j}\right)} = u(\gamma_i) \cdot u\left(d'_{i,j}\right).$$
(17)

Subsequently, we derive the variance of the mean, including the non-vanishing covariances:

$$u^2\left(\langle d'\rangle_j\right) \approx \sum_{i=1}^{3} \kappa_i^2 \cdot u^2\left(d'_{i,j}\right)$$

$$+ \left(\kappa_3^2 \cdot u^2(\gamma_1) + \kappa_2^2 \cdot u^2(\gamma_2) + \kappa_1^2 \cdot u^2(\gamma_3)\right)$$

$$\cdot \left(\frac{d'_{3,j} - d'_{1,j}}{\gamma_3 - \gamma_1}\right)^2 + 2\left(\kappa_1\kappa_3 \cdot u\left(d'_{1,j}\right) \cdot u(\gamma_1)\right.$$

$$- \kappa_2^2 \cdot u\left(d'_{2,j}\right) \cdot u(\gamma_2)$$

$$\left. + \kappa_1\kappa_3 \cdot u\left(d'_{3,j}\right) \cdot u(\gamma_3)\right)$$

$$\cdot \left(\frac{d'_{3,j} - d'_{1,j}}{\gamma_3 - \gamma_1}\right).$$
(18)

The uncertainty of $\langle d'\rangle_w$ can be obtained analogously to that of $\langle d'\rangle_j$ following GUM[41]; however, the uncertainties of the weights (λ_j) are highly correlated, as is evident from their definition in Equation (15). We decided to use the Monte Carlo method according to Supplement 1 of GUM[42] to estimate the uncertainty of $\langle d'\rangle_w$ based on Equations (13) and (14). A probability-density function needs to be assigned to each input quantity ($\langle d'\rangle_j$, D_j, and c_j), which was performed according to Supplement 1 of GUM.[42] The assumption of a rectangular distribution or truncated normal distribution (i.e., excluding unrealistic zero or negative values) for D_j and c_j had no relevant impact on the results. The results obtained using the rectangular distribution are presented herein. A normal distribution with variance $u^2(\langle d'\rangle_j)$ (as given in Equation (18)) is assigned to the input quantity $\langle d'\rangle_j$. We used 10^6 histories, as this is expected to produce a 95% confidence interval correct for one or two significant digits.[42] The standard deviation of the distribution of $\langle d'\rangle_w$ from the Monte Carlo simulation was used as an estimation of the uncertainty of $\langle d'\rangle_w$.

Careful calculation of these uncertainties is critical for defining proper acceptance conditions with respect to a given threshold. For a possible type test involving just a lower threshold for $\langle d'\rangle_w$, the 90 % coverage interval is calculated from the standard uncertainty according to accepted guidelines.[43,44] Thus, the probability that the estimate is outside the acceptance interval is 5 % if the lower boundary of the 90 % coverage interval is greater than or equal to the tolerance limit.[43,45] The criterion for acceptance would then be that the lower limit of the 90 % coverage interval is at least equal to or above the acceptance threshold.

TABLE 1 Diameter and contrast (with 95% confidence intervals) of the four small-contrast rods of the MITA-body phantom

Contrast rods	D (mm)	c (HU)
1	3.0 ± 0.1	14.0 ± 0.5
2	5.0 ± 0.1	7.0 ± 0.5
3	7.0 ± 0.2	5.0 ± 0.5
4	10.0 ± 0.2	3.0 ± 0.5

The dose levels γ used to compute the dose efficiency $\langle d'\rangle_w$ are given by the weighted CT dose index ($CTDI_w$) that is displayed at every CT scanner. It is based on dose measurements on the respective system using a pencil-type ionization chamber to measure the dose at different positions in the standard CT dosimetry phantom for all available tube voltages and a standardized scan protocol. These scanner-specific dose values are then adjusted to other CT protocols (by means of the tube current, pitch, etc.) The International Atomic Energy Agency quantified the uncertainty of $CTDI_w$ values to be 7.2%.[46] In our study, the uncertainty was assumed to be 8%.

3 | MATERIALS AND METHODS

3.1 | Test phantom and CT imaging

To test the practicability of the proposed measurement protocol and computation of the FOM, CT images of the Medical Imaging Technology Alliance (MITA) body phantom (CCT189; The Phantom Lab, Salem, NY, USA) were acquired (Figure 1). This phantom was developed for low-contrast object detection to assess the dose-reduction potential of iterative reconstruction algorithms.[47] It consists of a homogeneous cylinder with a diameter of 20 cm and an axial length of 20 cm, containing two sets of contrast rods with lengths of 4 cm (smaller rods) and 1 cm (larger rods). Each small contrast rod had a corresponding larger rod with the same density and contrast intended for reliably measuring the contrast. The image analysis focused on the smaller rods, as listed in Table 1. The phantom was equipped with an additional ring-shaped attenuator (CTP 653-20; The Phantom Lab) with an outer diameter of 32 cm and axial length of 20 cm to better approximate the dimensions of a reference adult, thus realizing realistic attenuation conditions for the abdominal protocol. The phantom was attached to a wooden box with two hooks (Figure 1a). This ensured that the orientation of the phantom was always comparable and that the contrast rods had a constant angular orientation.
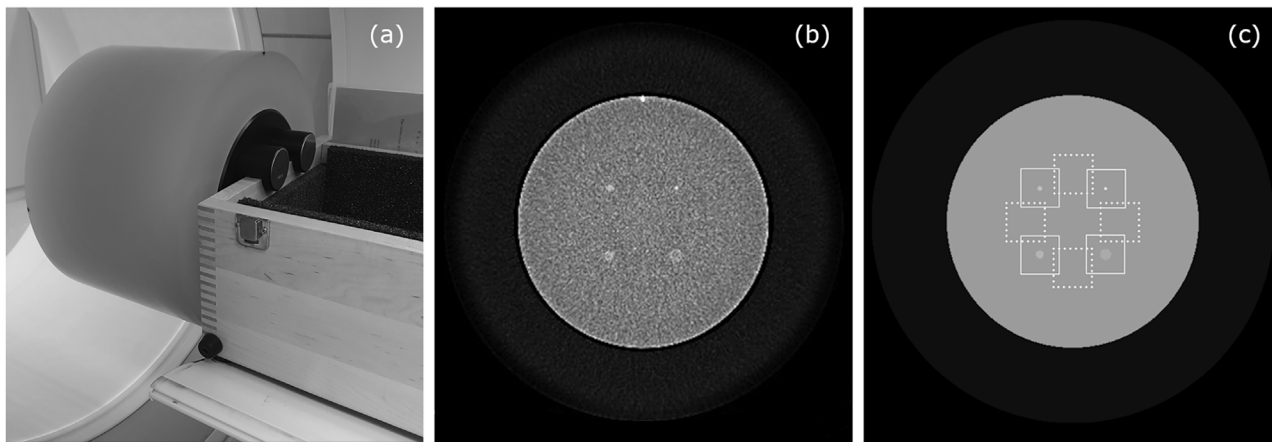
FIGURE 1 (a) Medical Imaging Technology Alliance (MITA) body phantom (dark) with a 32 cm-diameter attenuator (opaque) on the patient table. (b) Computed tomography (CT) image acquired at a very high dose level (weighted CT dose index (CTDI$_w$): 87 mGy) for illustration. (c) Schematic illustration of the MITA-body phantom. Solid squares correspond to regions of interest (ROIs) with contrast objects, and dotted squares correspond to ROIs without contrast objects (noise).

Measurements were performed using three single-source CT systems from different manufacturers. They represent different technical levels and generations and are subsequently referred to as "CT 1" (64 detector rows), "CT 2" (128 detector rows), and "CT 3" (64 detector rows). Two different phantom specimens were used: one was imaged on "CT 2" and the other on "CT 1" and "CT 3." The algorithms used for image reconstruction are classified into three groups: traditional filtered-back projection (FBP), hybrid/statistical iterative reconstruction (H/SIR), and full/partial/advanced model-based iterative reconstruction (MBIR), following the convention of Greffier et al.[48]

The standardized protocol used for CT measurements simulates the examination of low-contrast objects in the abdomen (e.g., liver metastases) with a high resolution. The acquisition and reconstruction settings are presented in Table 2. The small differences between the acquisition parameters used for the three devices were due to the technical constraints. Scanning was performed for all three $CTDI_w$ values. For each dose level, 200 phantom scans were obtained. Image data were evaluated from two adjacent slices (5 mm apart; referred to as slice "1" and "2") of the same CT scan of the phantom. Although the data were acquired in the axial rather than spiral mode, there is some correlation of the noise pattern of the two reconstructed CT datasets. Nevertheless, the use of these pseudo-independent datasets allows a rudimentary assessment of reproducibility.

3.2 | Software implementations and image analysis

Two separate implementations of the task-based observer were developed, one in MATLAB (R2020a;

The MathWorks Inc., Natick, USA) by the German National Metrology Institute (Physikalisch-Technische Bundesanstalt, PTB) and one in Python (Version 3.6.8, Python Software Foundation, https://www.python.org/) by the Federal Office for Radiation Protection (Bundesamt für Strahlenschutz, BfS). These are referred to as implementations A and B. The significance of differences in the central tendency between the implementations was tested by the Wilcoxon signed rank test that does not require assuming normality and equal variance. Statistical analysis was performed using SigmaPlot (Version 13.0; Systat Software GmbH, Erkrath, Germany). A p value of 0.05 was used as threshold for significance.

Eight regions of interest (ROIs) were defined on each image: four centered on small contrast rods (solid squares in Figure 1), and four in homogeneous regions (dotted squares in Figure 1). The square ROIs had a size of 64×64 pixels. In total, 200 signal-present and 200 signal-absent ROIs were analyzed for each $CTDI_w$ value, contrast rod, and slice position.

To define the square ROIs required for the model observer analyses, we used the following approaches: In the first step, the average of the 200 CT images acquired from a slice at the highest dose level was computed. The second step differed in how approximate positions of the four rods were derived from this image: Semi-automatic (implementation A) by manually selecting the four rods or fully automatic (implementation B) by segmenting the outer contour of the phantom (and thus its center) and then defining the position of the rods based on the manufacturer's specification of the used MITA phantom. In the third step, circular templates with the nominal diameters of the rods were separately placed at the position of the approximatively determined ROIs and then exactly matched by a cross-correlation approach to the rods

TABLE 2 Protocol settings for data acquisition and image reconstruction

CT system	CT 1	CT 2	CT 3
Parameter	Setting/value		
Data acquisition			
Protocol	Abdomen (adults)	Abdomen (adults)	Abdomen (adults)
Scan mode	Axial	Axial	Axial
Collimation (mm)	64×0.625	128×0.6	64×0.625
Rotation time (s)	0.5	0.5	0.5
Tube voltage (kV)	120	120	120
Beam filter	Default[a]	Default[a]	Default[a]
$CTDI_w$ (mGy)	2.0, 4.8, 12.0	1.7, 3.8, 10.1	1.8, 4.5, 10.8
Image reconstruction			
Reconstructed slice thickness (mm)	5	5	5
Convolution kernel	Default[a]	Default[a]	Default[a]
Reconstruction field of view (mm)	250	250	250
Reconstruction algorithm	FBP and H/SIR[b]	FBP and MBIR[b]	FBP, H/SIR[b] and MBIR[b]
Matrix size	512×512	512×512	512×512

[a]Routine body protocols.
[b]Maximum value/strength available in clinical practice.

visualized on the average CT image. The center coordinates of the matched templates were used in the final step to place the square ROIs as shown in Figure 1(c).

ROIs were defined only on the images acquired at the highest dose level and were then transferred to that acquired at lower doses. The signal-absent ROIs were positioned on the same images in homogeneous regions between the rods at the same distance from the phantom center using a 45° rotation.[13]

To test the effects of the two algorithms used for ROI selection and of the different implementations of the same model observer on the computed FOM separately, the ROIs from implementation A were additionally used to estimate the FOM using implementation B of the model observer.

4 | RESULTS

The distances from the center of the MITA phantom to the center of the rods varied by up to 6% (5 mm) compared to the nominal distance, as declared by the manufacturer. As both implementations (A and B) do not utilize a fixed template of the whole phantom, but rather localize each rod individually, they were both able to detect the actual positions of the rods.

Figure 2 shows representative axial CT images of the MITA-body phantom to illustrate the different noise textures yielded by three reconstruction algorithms (FBP, H/SIR, and MBIR). As an example, d' values are plotted in Figure 3 for measurements performed at "CT 3" as a function of the applied dose ($CTDI_w$) for the four con-

trast rods and three different reconstruction algorithms (FBP, H/SIR, and MBIR). Despite the appreciable variation of the d' values (1.3–5.5), the systematic effect of the dose and reconstruction algorithm was clearly visible. Whereas image reconstruction with the FBP and H/SIR algorithms yielded almost identical results, the MBIR algorithm substantially increased detectability. Because detectability d' depends on various factors, it is not well suited as an FOM for comparing CT systems with respect to their dose efficiency. As Figures 4 and 5 demonstrate, a more comprehensive and robust characterization is obtained when the detectabilities are averaged over the different dose levels $\langle d' \rangle_j$ (cf. Equation (13)), and finally, over the four rods $\langle d' \rangle_w$ (cf. Equation (14)).

The relative differences between the results obtained using both implementations were up to 24% for d', 7.1% for $\langle d' \rangle$, and 2.4% for $\langle d' \rangle_w$. The difference between the ROI positions from implementation A and B are up to four pixels for the signal-present ROIs (mean: 0.8 pixels). When identical ROIs were used (in this case, the ROIs obtained by implementation A), the results obtained by both implementations of the model observer were nearly identical, with relative deviations of less than 0.1% for d', 0.03% for $\langle d' \rangle$, and 0.02% for $\langle d' \rangle_w$, which underlines the relevance of the ROI definitions.

As mentioned in Section 3.1, CT images from pseudo-independent slices ("1" and "2") were separately analyzed (the results from the second dataset (slice "2") are not included in the figures for reasons of clarity). The analysis of both datasets ("1" and "2") yielded comparable results, with a mean difference of 0.45 % (−35% to
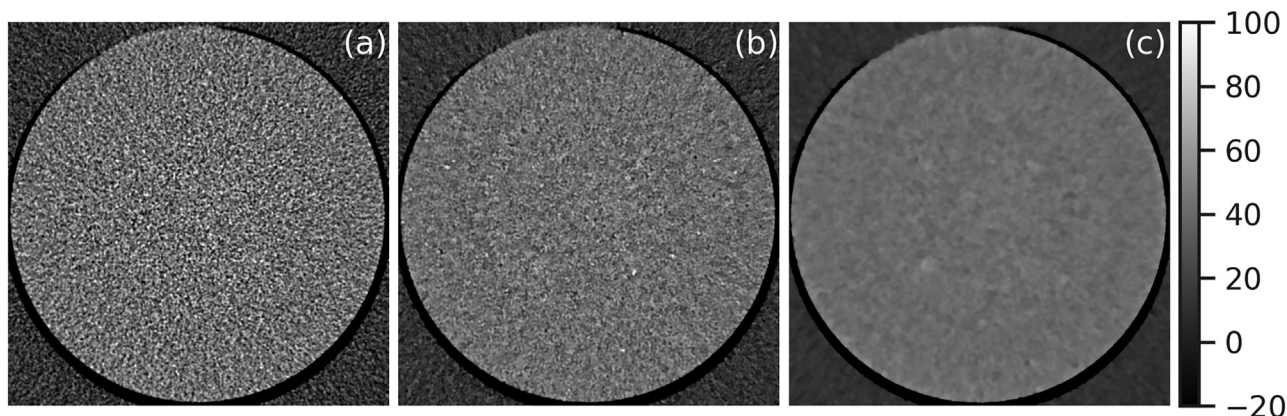
FIGURE 2 Axial image of the MITA-body phantom reconstructed with the (a) traditional filtered-back projection (FBP), (b) hybrid/statistical iterative reconstruction (H/SIR), and (c) full/partial/advanced model-based iterative reconstruction (MBIR) algorithms from raw data acquired on "CT 3" ($CTDI_w$: 11 mGy) demonstrating the different noise textures. Note that the low-contrast inserts are barely visible in these low-dose images. Nevertheless, the model observer was able to calculate the detectability of the inserts.
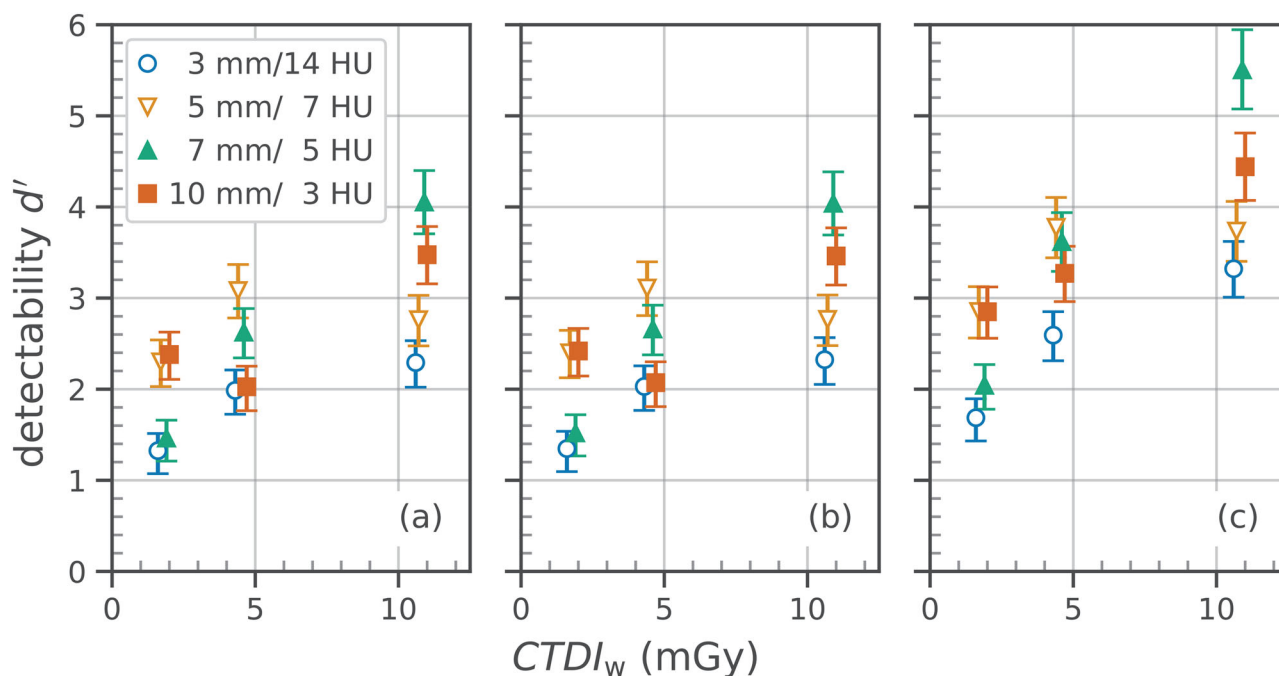
FIGURE 3 Detectabilities d' computed from CT images reconstructed with the (a) FBP, (b) H/SIR, and (c) MBIR algorithms from raw data acquired at "CT 3" at different dose levels. d' values for the four contrast rods are plotted as symbols, and the error bars represent a confidence level of 95%.

32%) for d', 0.64% (−17% to 13%) for $\langle d' \rangle$, and 0.81% (−0.15% to 1.9%) for $\langle d' \rangle_w$.

5 | DISCUSSION

The presented measurement and data analysis procedures enable robust and reliable assessment of the quality of CT images with low-contrast structures in relation to the radiation dose applied for their acquisition. All computational steps included an assessment of the related uncertainties by adapting the concepts

presented by Wunderlich[39] and the Joint Committee for Guides in Metrology.[41,42] The quantity $\langle d' \rangle_w$ is well suited not only for quantifying the effect of different reconstruction algorithms, but also for the characterization of the dose efficiency of CT systems, and can thus lay the foundation for possible standardized type testing procedures. CHO without internal noise is appropriate for technical image-quality assessment because it approximates HO (which is the ideal linear observer).

The time required for the proposed methodology was acceptable for a type testing procedure. With some experience, data acquisition can be performed in less
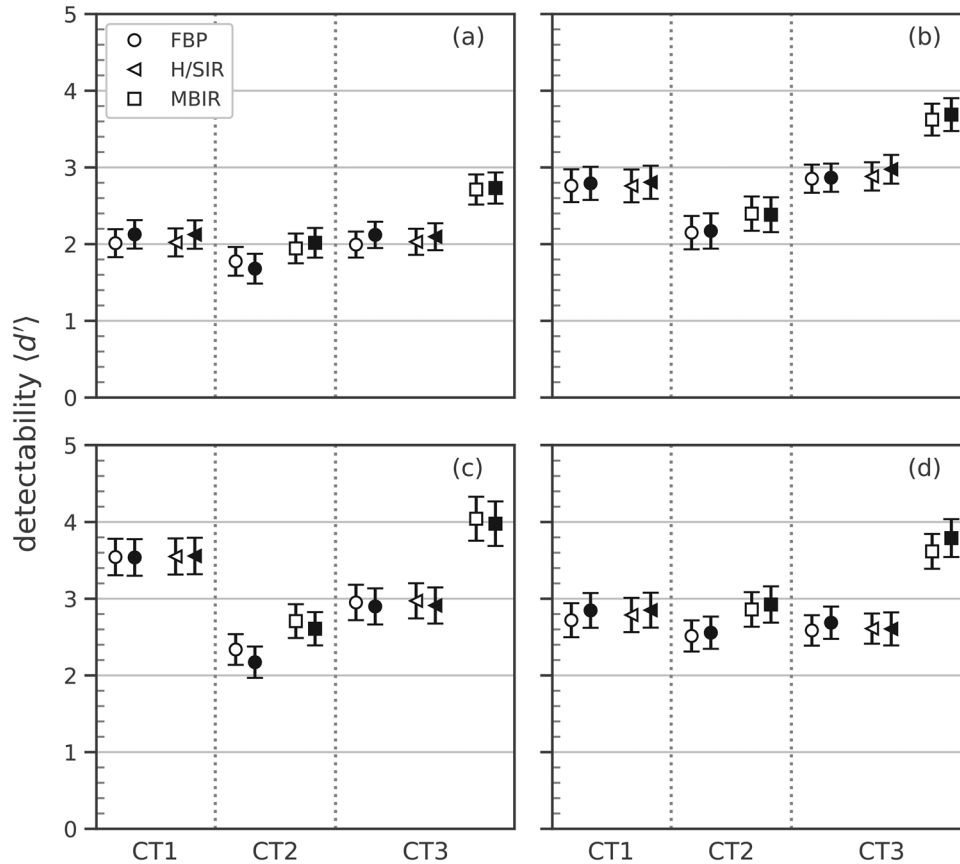
FIGURE 4 Mean detectabilities $\langle d' \rangle$ computed according to Equation (13) from CT images of the MITA phantom acquired from three different CT systems at three dose levels (cf. Table 2). Image reconstruction was performed using the FBP, H/SIR, and MBIR algorithms. Statistical image analysis was performed using implementations A (open symbols) and B (filled symbols). The diameters/contrasts of the four rods were (a) 3 mm/14 HU, (b) 5 mm/7 HU, (c) 7 mm/5 HU, and (d) 10 mm/3 HU. Error bars indicate the 95% coverage interval. Differences between implementations A and B were statistically significant for (b), (c), and (d). It should be noted that $\langle d' \rangle$ depends on both the diameter and contrast of the rods.
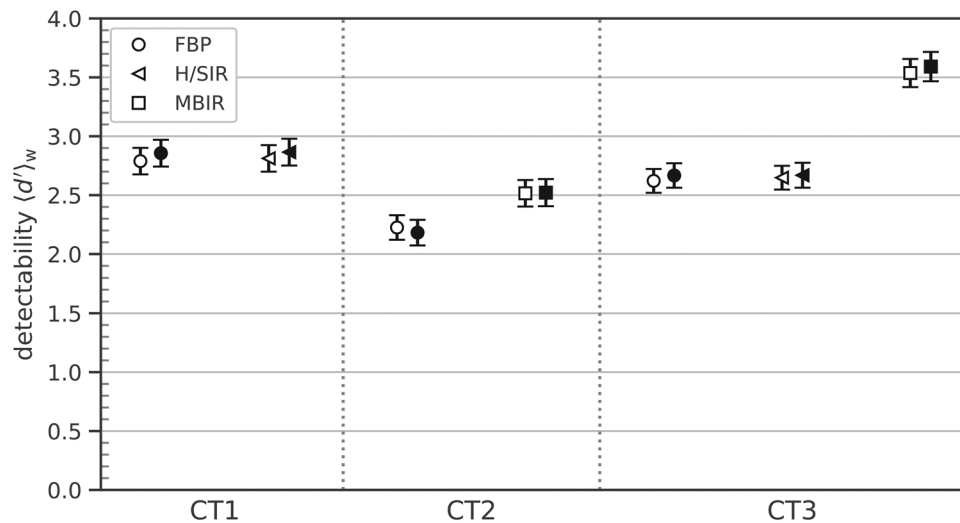

FIGURE 5 Weighted detectabilities $\langle d' \rangle_w$ computed according to Equation (14) from CT images acquired from the MITA phantom using three different CT systems and four dose levels (cf. Table 2). Statistical image analysis was performed with implementations A (open symbols) and B (filled symbols). Error bars indicate the 95 % coverage interval. Differences between implementations A and B were not statistically significant.

than 2 h. The time demand for image sorting, processing, and analysis can be reduced using in-house made or available software (e.g., Hernandez-Giron et al.,[40] Wunderlich,[36] or the tools developed in the present study) to automate otherwise time-consuming steps.

When evaluating identical ROIs, the results obtained using the two implementations of the model observer showed excellent agreement. However, even slight inaccuracies in the definition of the ROIs can substantially affect the estimated FOM, as revealed by the different results obtained by the two implementations (A and B) utilized in the present study. Neither method used for ROI definition was affected by inter-phantom variabilities caused by manufacturing tolerances in the radial distance of the contrast rods.

We also checked our implementations of the model observer by applying it to a dataset used in an international intercomparison exercise.[19] For this intercomparison, Ba et al. acquired CT images of a homogeneous phantom with a diameter of 20 cm at a $CTDI_{\mathrm{w}}$ of 15 mGy, and added virtual lesions with diameters of 6, 8, and 10 mm and a contrast of approximately 4 HU to the images. All study participants received the same image set with signal-present and signal-absent ROIs and used the CHO observer with DoG channels, as proposed by Abbey and Barrett.[21] By analyzing these ROIs using our software, we achieved d' values of 4.64, 5.81, and 7.56 for the three lesions. This corresponds very well to the median values of 4.71, 5.93, and 7.76 reported by Ba et al. The variations between the participants in the intercomparison were less than 5%, 16%, and 18% for the three lesions. Although the measurement conditions in the present study do not exactly match those of Ba et al. (lower dose levels and larger phantom size in our study), it can be concluded that the magnitude of the reported variations between our two implementations of less than 0.1% is much smaller than that in the intercomparison.

In combination with the specified scan protocol, the proposed FOM $\langle d' \rangle_{\mathrm{w}}$ is well suited for use as a dose-efficiency index, enabling the benchmarking of different CT scanners and reconstruction algorithms. For type testing, an acceptance threshold or tolerance limit for $\langle d' \rangle_{\mathrm{w}}$ must be defined (Section 2.5). The acceptance condition is satisfied if the lower bound of the 90% confidence interval of the determined weighted detectability, $\langle d' \rangle_{\mathrm{w}}$, equals or exceeds this threshold.

To define a reasonable threshold for type testing, an extensive measurement campaign involving various CT systems from different manufacturers is planned. This campaign will also allow us to identify possible problems and limitations of the implemented model observer and software. Part of this study will be a robustness check with several repeated measurements using the same CT system as well as measurements for different systems of the same CT type to assess intra- and inter-system variabilities. Moreover, we will investigate in an upcoming study the effect of variations in the contrast of the four rods in the MITA phantom, as well as the use of more challenging phantoms, such as the MITA-head phantom (CCT191; The Phantom Lab). The aim is to establish a robust and reliable model-observer approach that can be defined as part of mandatory type testing by CT manufacturers.

However, it must be noted that the proposed approach for benchmarking CT systems is based on images of a standardized but simple phantom. In this regard, a problem could arise when innovative reconstruction algorithms, especially those using deep-learning algorithms, are used that might "learn" the structure of the phantom that is regularly used for quality assurance, and thus yield an unrealistically high dose-efficiency index. To address this challenge, it is essential to precisely specify quality assurance protocols for image reconstruction and post-processing.

6 | CONCLUSION

The presented automatable software approach yields an objective FOM and is thus a good candidate for robust and reliable benchmarking of different CT systems and reconstruction algorithms. An essential advantage of the proposed model observer is that FOM uncertainties can be provided, which is an indispensable prerequisite for type testing. However, it is not only necessary to standardize the model observer, but also to use largely automated and standardized image-processing software. In particular, this applies to the geometrical definition of the ROIs containing the contrast objects, such that inaccuracies in this processing step can be reduced. Using the presented and evaluated model observer and post-processing tools for quality assurance, cumbersome ROC analyses with human observers, which are affected by high inter-individual variations, can be replaced by an objective and automated approach.

CONFLICT OF INTEREST STATEMENT

The authors have no relevant conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Maximilian Göppel
https://orcid.org/0009-0006-1161-4317

REFERENCES

1. United Nations Scientific Committee on the Effects of Atomic Radiation. UNSCEAR 2020/2021 Report: "Sources, effects and risks of ionizing radiation": Annex A: Evaluation of medical exposure to ionizing radiation. 2022. https://www.unscear.org/unscear/en/publications/2020_2021_1.html
2. Singh S, Kalra MK, Do S, et al. Comparison of hybrid and pure iterative reconstruction techniques with conventional filtered back projection: dose reduction potential in the abdomen. *J Comput Assisted Tomogr*. 2012;36(3):347-353. doi:10.1097/RCT.0b013e31824e639e
3. Mohammadinejad P, Mileto A, Yu L, et al. CT noise-reduction methods for lower-dose scanning: strengths and weaknesses of iterative reconstruction algorithms and new techniques. *Radiographics*. 2021;41(5):1493-1508. doi:10.1148/rg.2021200196
4. Lell MM, Kachelrieß M. Recent and upcoming technological developments in computed tomography: high speed, low dose, deep learning, multienergy. *Invest Radiol*. 2020;55(1):8-19. doi:10.1097/rli.0000000000000601
5. Hsieh J, Flohr T. Computed tomography recent history and future perspectives. *J Medical Imag (Bellingham, Wash)*. 2021;8(5):052109. doi:10.1117/1.Jmi.8.5.052109
6. ICRP. The 2007 recommendations of the International Commission on Radiological Protection: ICRP publication 103. *Ann ICRP*. 2007;(37.2007,2/4):141. doi:10.1016/j.icrp.2007.10.003
7. Riederer SJ, Pelc NJ, Chesler DA. The noise power spectrum in computed X-ray tomography. *Phys Med Biol*. 1978;23(3):446-454. doi:10.1088/0031-9155/23/3/008
8. Brooks RA, Di Chiro G. Statistical limitations in x-ray reconstructive tomography. *Med Phys*. 1976;3(4):237-240. doi:10.1118/1.594240
9. Barrett HH, Myers KJ. *Foundations of image science*. Wiley-Interscience; 2004.
10. 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(7):4115-4122. doi:10.1118/1.4725171
11. Illers H, Buhr E, Hoeschen C. Measurement of the detective quantum efficiency (DQE) of digital X-ray detectors according to the novel standard IEC 62220–1. *Radiat Protect Dosimetry*. 2005;114(1-3):39-44. doi:10.1093/rpd/nch507
12. Sharp P, Barber DC, Brown DG, et al. Medical imaging: The assessment of image quality. ICRU Report; 1996:88.
13. 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(7):071904. doi:10.1118/1.4881148
14. International Commission on Radiation Units & Measurements. ICRU Report No. 87: radiation dose and image-quality assessment in computed tomography. *J ICRU*. 2012;12(1):1-149. doi:10.1093/jicru/ndt007

15. Samei E, Bakalyar D, Boedeker K, et al. *Performance evaluation of computed tomography systems - The report of AAPM Task Group 233*. AAPM; 2019.
16. 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(2):R1-75. doi:10.1088/0031-9155/60/2/R1
17. Kundel HL, Berbaum KS, Dorfman DD, Gur D, Metz CE, Swensson RG. *Receiver operating characteristic analysis in medical imaging*. Oxford Univ. Press; 2008:62. ICRU Report.
18. Racine D, Ryckx N, Ba A, et al. Task-based quantification of image quality using a model observer in abdominal CT: a multicentre study. *Eur Radiol*. 2018;28(12):5203-5210. doi:10.1007/s00330-018-5518-8
19. Ba A, Abbey CK, Baek J, et al. Inter-laboratory comparison of channelized hotelling observer computation. *Med Phys*. 2018;45(7):3019-3030. doi:10.1002/mp.12940
20. International Electrotechnical Commission. Evaluation and routine testing in medical imaging departments - Part 3–5: Acceptance and constancy tests - Imaging performance of computed tomography X-ray equipment. IEC 61223-3-5:2019. 2.0 ed2019.
21. Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. *J Opt Soc Am A, Opt, Image Sci Vision*. 2001;18(3):473-488. doi:10.1364/josaa.18.000473
22. Verdun FR, Racine D, Ott JG, et al. Image quality in CT: from physical measurements to model observers. *Phys Med*. 2015;31(8):823-843. doi:10.1016/j.ejmp.2015.08.007 . Physica Medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB).
23. Racine D, Ba AH, Ott JG, Bochud FO, Verdun FR. Objective assessment of low contrast detectability in computed tomography with channelized hotelling observer. *Phys Med*. 2016;32(1):76-83. doi:10.1016/j.ejmp.2015.09.011 . Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB).
24. Burgess AE. Visual perception studies and observer models in medical imaging. *Semin Nucl Med*. 2011;41(6):419-436. doi:10.1053/j.semnuclmed.2011.06.005
25. Diaz I, Abbey CK, Timberg PAS, et al. Derivation of an observer model adapted to irregular signals based on convolution channels. *IEEE Trans Med Imag*. 2015;34(7):1428-1435. doi:10.1109/tmi.2015.2395433
26. Barrett HH, Abbey CK, Clarkson E. Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions. *J Opt Soc Am A, Opt, Image Sci Vision*. 1998;15(6):1520-1535. doi:10.1364/josaa.15.001520
27. Soares EJ. *Attenuation, noise, and image quality in single-photon emission computed tomography*. University of Arizona; 1994. Accessed 2011-10-31t18:21:11z. http://hdl.handle.net/10150/186834
28. Smith WE, Barrett HH. Hotelling trace criterion as a figure of merit for the optimization of imaging systems. *J Opt Soc Am A*. 1986;3(5):717. doi:10.1364/josaa.3.000717
29. Fiete RD, Barrett HH, Smith WE, Myers KJ. Hotelling trace criterion and its correlation with human-observer performance. *J Opt Soc Am A, Opt Image Sci*. 1987;4(5):945-953. doi:10.1364/josaa.4.000945
30. Hotelling H. The generalization of student's ratio. *Ann Math Statist*. 1931;2(3):360-378. doi:10.1214/aoms/1177732979
31. Eckstein MP, Abbey CK, Bochud FO. A practical guide to model observers for visual detection in synthetic and natural noisy images. In: Kundel HL, Beutel J, van Metter RL, eds. *Handbook of Medical Imaging*. SPIE. SPIE Press monograph; 2000:593-628.

32. Anton M, Veldkamp WJH, Hernandez-Giron I, Elster C. RDI-a regression detectability index for quality assurance in: x-ray imaging. *Phys Med Biol*. 2020;65(8). doi:10.1088/1361-6560/ab7b2e . ARTN 085017.

33. Abbey CK. *Assessment of reconstructed images*. The University of Arizona; 1998. http://hdl.handle.net/10150/288785

34. Abbey C, Bochud F. In: Kundel HL, Beutel J, van Metter RL, eds. *Modeling visual detection tasks in correlated image noise with linear model observers*. SPIE Press monograph; 2000:629-654. Handbook of medical imaging. SPIE.

35. Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. *J Opt Soc Am A, Opt Image Sci*. 1987;4(12):2447-2457. doi:10.1364/josaa.4.002447

36. Wunderlich A, IQmodelo: Statistical Software for Task-Based Image Quality Assessment with Model (or Human) Observers. Updated 15.04.2016. Accessed 16.03.2022, https://github.com/DIDSR/IQmodelo

37. Chawla AS, Samei E, Saunders R, Abbey C, Delong D. Effect of dose reduction on the detection of mammographic lesions: a mathematical observer model analysis. *Med Phys*. 2007;34(8):3385-3398. doi:10.1118/1.2756607

38. Hernandez-Giron I, Calzado A, Geleijns J, Joemai RMS, Veldkamp WJH. Comparison between human and model observer performance in low-contrast detection tasks in CT images: application to images reconstructed with filtered back projection and iterative algorithms. *Brit J Radiol*. 2014;87(1039):20140014. doi:10.1259/bjr.20140014

39. Wunderlich A, Noo F, Gallas BD, Heilbrun ME. Exact confidence intervals for channelized Hotelling observer performance in image quality studies. *IEEE Trans Med Imaging*. 2015;34(2):453-464. doi:10.1109/TMI.2014.2360496

40. Hernandez-Giron I, Geleijns J, Calzado A, Veldkamp WJH. Automated assessment of low contrast sensitivity for CT systems using a model observer. *Med Phys*. 2011;38(Suppl 1):S25. doi:10.1118/1.3577757

41. Joint Committee for Guides in Metrology. *Evaluation of measurement data: Guide to the expression of uncertainty in measurement*; 2008.

42. Joint Committee for Guides in Metrology. *Evaluation of measurement data: Supplement 1 to the GUM - Propagation of distributions using a Monte Carlo method*; 2008.

43. Joint Committee for Guides in Metrology. *Evaluation of measurement data: The role of measurement uncertainty in conformity assessment*; 2012.

44. Deutsches Institut für Normung e. V. Grundlagen der Meßtechnik - Teil 3: auswertung von Messungen einer einzelnen Meßgröße, Meßunsicherheit: auswertung von Messungen einer einzelnen Meßgröße. *Meßunsicherheit*. 1319-3:1996-05. Berlin: Beuth; 1996.

45. Strahlenschutzkommission. Methodik zur Berücksichtigung von Messunsicherheiten bei messtechnischen Prüfungen im Geltungsbereich der Röntgenverordnung und der Strahlenschutzverordnung: empfehlung der Strahlenschutzkommission. *Bonn*; 2016.

46. Pernicka F, McLean ID. Dosimetry in diagnostic radiology: an international code of practice. Technical reports series. *Int Atom Energy Agency*. 2007;1:359. online resource.

47. CT Group of the X-Ray Imaging Section of the Medical Imaging Technology Alliance. Computed tomography image quality (CTIQ): Low-contrast detectability (LCD) assessment when using dose reduction technology. 2017. https://www.nema.org/Standards/view/Computed-Tomography-Image-Quality-CTIQ-Low-Contrast-Detectability-LCD-Assessment-When-Using-Dose-Reduction-Technology

48. Greffier J, Frandon J, Larbi A, Beregi JP, Pereira F. CT iterative reconstruction algorithms: a task-based image quality assessment. *Eur Radiol*. 2020;30(1):487-500. doi:10.1007/s00330-019-06359-6