Original paper

Effects of various generations of iterative CT reconstruction algorithms on low-contrast detectability as a function of the effective abdominal diameter: A quantitative task-based phantom study

Anais Viry[a,*], Christoph Aberle[b], Damien Racine[a], Jean-François Knebel[c,d], Sebastian T. Schindera[e], Sabine Schmidt[f], Fabio Becce[f,1], Francis R. Verdun[a,1]

[a] *Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland*
[b] *Department of Radiology, University Hospital Basel, University of Basel, Basel, Switzerland*
[c] *Laboratory for Investigative Neurophysiology (The LINE), Department of Radiology and Department of Clinical Neurosciences, Lausanne University Hospital, Lausanne, Switzerland*
[d] *Electroencephalography Brain Mapping Core, Center for Biomedical Imaging (CIBM), Lausanne and Geneva, Switzerland*
[e] *Department of Radiology, Kantonsspital Aarau, Aarau, Switzerland*
[f] *Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Lausanne, Switzerland*

ARTICLE INFO

ABSTRACT

Purpose: To investigate how various generations of iterative reconstruction (IR) algorithms impact low-contrast detectability (LCD) in abdominal computed tomography (CT) for different patient effective diameters, using a quantitative task-based approach.

Methods: Investigations were performed using an anthropomorphic abdominal phantom with two optional additional rings to simulate varying patient effective diameters (25, 30, and 35 cm), and containing multiple spherical targets (5, 6, and 8 mm in diameter) with a 20-HU contrast difference. The phantom was scanned using routine abdominal protocols ($CTDI_{vol}$, 5.9–16 mGy) on four CT systems from two manufacturers. Images were reconstructed using both filtered back-projection (FBP) and various IR algorithms: ASiR 50%, SAFIRE 3 (both statistical IRs), ASiR-V 50%, ADMIRE 3 (both partial model-based IRs), or Veo (full model-based IR). Section thickness/interval was 2/1 mm or 2.5/1.25 mm, except 0.625/0.625 mm for Veo. We assessed LCD using a channelized Hotelling observer with 10 dense differences of Gaussian channels, with the area under the receiver operating characteristic curve (AUC) as a figure of merit.

Results: For the smallest phantom (25-cm diameter) and smallest lesion size (5-mm diameter), AUC for FBP and the various IR algorithms did not significantly differ for any of the tested CT systems. For the largest phantom (35-cm diameter), Veo yielded the highest AUC improvement (8.5%). Statistical and partial model-based IR algorithms did not significantly improve LCD.

Conclusion: In abdominal CT, switching from FBP to IR algorithms offers limited possibilities for achieving significant dose reductions while ensuring a constant objective LCD.

1. Introduction

Over the last twenty years, computed tomography (CT) examinations have been performed with increasing frequency in Western countries, leading to improved patient care. This has, however, substantially increased the impact of CT on the collective effective radiation dose [1].

Automatic tube current modulation (ATCM) has first been proposed to reduce radiation exposure by accounting for the patient's X-ray attenuation. This is one of the most effective tools for optimizing patient exposure, particularly for thoracic and abdominal CT examinations. One of the ATCM strategies, developed by Siemens, is to automatically adapt the tube current level depending on the patient body habitus. Another strategy, used by GE, is to manually set a noise index (NI) level [2,3]. In all cases, radiologists must tolerate higher noise levels in larger patients because of the exponential nature of X-ray attenuation [4,5]. To offer further dose reductions, CT manufacturers have subsequently replaced the traditional filtered back-projection

(FBP) reconstruction algorithm with iterative reconstruction (IR) algorithms, which can be distinguished into two generations: statistical or hybrid IR, and model-based IR [6]. Model-based IR algorithms include the modeling of physical and optical aspects in the reconstruction process. The modeling complexity impacts the image reconstruction time, such that manufacturers can propose either a fast partial version or a longer full version of their model-based IR algorithm.

When aiming at reducing radiation dose, the optimization process of CT protocols should ensure that the clinical question can still always be answered, while keeping the radiation exposure as low as reasonably achievable [7]. Among all clinical data contained in patient images, one of the most challenging task is the detection of low-contrast lesions, such as in non-contrast CT scans of the abdomen, particularly in large patients. Hence, it appears necessary to assess the low-contrast detectability (LCD) performances of various IR algorithms with the routine clinical CT settings for different patient body habitus before implementing drastic dose reductions with IR [8].

To assess the impact of dose reductions on LCD, one robust and objective approach involves the use of anthropomorphic CT phantoms with simulated low-contrast lesions. Previous studies have assessed the potential improvement of LCD with IR algorithms for one single phantom size, and often applied subjective and/or objective image quality parameters based on the technical efficacy (image noise, contrast-to-noise ratio (CNR), and spatial resolution) [9–11]. However, since it is difficult to correlate these image quality metrics with an accurate assessment of diagnostic performance, it is of paramount importance to use image quality criteria linked to a specific and clinically relevant task, such as LCD, for different phantom sizes [12,13].

In this study, we used a task-based quantitative image quality assessment approach to investigate how the LCD for various patient effective diameters was impacted by the use of various generations of IR algorithms from two CT manufacturers, using routine clinical settings.

2. Materials and methods

2.1. Phantom

A commercially-available tissue equivalent anthropomorphic abdominal phantom (QRM, Moehrendorf, Germany) was used to mimic data obtained from non-contrast CT scans of patients. The phantom contains a 10-cm-diameter central hole into which modules can be inserted to assess image quality. We used a cylindrical module containing low-contrast spheres that mimic hypodense abdominal lesions, and a homogeneous module comprising the same tissue equivalent material as the phantom. The module with hypodense abdominal lesions contained 24 spheres of 8, 6, 5, 4, and 3 mm in diameter, with a 20-HU contrast difference compared to the background material in a single axial section. This contrast difference was obtained using plastic materials of low effective Z numbers and different mass densities. The effective diameter of this phantom is 25 cm, thus simulating the X-ray attenuation of a thin patient weighing ~50 kg (Fig. 1a). To simulate different patient body habitus, additional rings of soft tissue equivalent

were added around the core of the phantom. Using 2.5- and 5-cm-thick rings, respectively, we achieved effective diameters of 30 and 35 cm, thus simulating patients weighing ~75 (Fig. 1b) and ~100 kg (Fig. 1c), respectively. The correlation between the phantom effective diameter and patient weight was previously obtained by performing a linear regression between the effective abdominal diameter and patient weight for 500 patients.

2.2. CT protocols

We scanned the phantom both with and without its additional rings using four different CT systems: the SOMATOM Definition Flash and SOMATOM Definition Edge (Siemens Healthineers, Forchheim, Germany), and the Discovery CT750 HD and Revolution (GE Healthcare, Milwaukee, WI, USA). Before data acquisition, we selected on each CT system the routine protocols used for the following clinical task: detection of low-contrast liver lesions. The portovenous acquisition phase was considered. The ATCM settings were those used in routine clinical practice: the quality reference mAs (with the "average" curve strength) were 150 and 130 mAs for the Flash and Edge Siemens systems, respectively; while for both GE systems, NI values were set to 20, 26, and 32 HU, for the small, medium, and large phantoms, respectively. The $CTDI_{vol}$ were retrieved directly from the CTDI dose report. With these settings, the $CTDI_{vol}$ for all CT systems varied between 5.9 and 7.8 mGy for the small phantom, 8.5 and 11.6 mGy for the medium phantom, and 11.1 and 16 mGy for the large phantom. To further simulate clinical practice, the display FOV was adapted for each phantom size, thus yielding different pixel sizes. Data acquisition and image reconstruction parameters are summarized in Table 1:

These settings produced a total of 27 datasets, including three phantom sizes and two image reconstruction algorithms for three of the four tested CT systems, and three phantom sizes and three image reconstruction algorithms for one CT scanner.

To achieve statistical robustness with the utilized image quality metrics, each phantom with the module containing low-contrast spheres was scanned 15 times without moving the phantom between acquisitions. Then, the low-contrast module was replaced with its homogeneous counterpart, and the phantom was scanned three times using the exact same settings, to provide the required number of images.

2.3. Task-based image quality assessment

We used a task-based approach to quantitatively assess the image quality of routine abdominal CT protocols, noting that the detection of low-contrast lesions is the most challenging task for such protocols. LCD was assessed for each image reconstruction algorithm and phantom size for three lesion sizes that were considered as clinically relevant for the investigated protocols (8, 6, and 5 mm in diameter). In clinical practice, smaller focal liver lesions (< 5 mm in diameter) may not be reliably characterized with the routine CT settings due to the poor low-contrast resolution of CT [14].
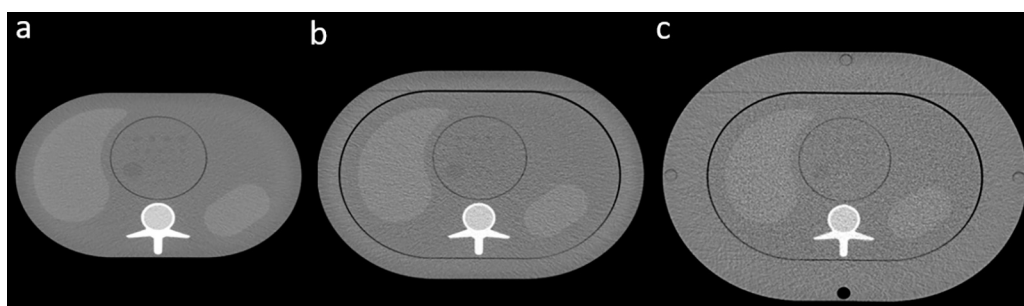

Fig. 1. Axial CT images of the anthropomorphic abdominal phantom simulating patients with weights of 50 (a), 75 (b), and 100 kg (c), respectively.

Table 1
Summary of the utilized CT data acquisition and image reconstruction parameters.

	Siemens Flash	Siemens Edge	GE Discovery CT750 HD	GE Revolution
Data acquisition parameters				
Tube potential (kVp)	120	120	120	120
Gantry revolution time (s)	0.5	0.5	0.6	0.5
Beam collimation (mm)	38.4	38.4	40	40
Pitch	0.6	0.8	0.986	0.975
ATCM settings[1]	Reference mAs = 150 mAs	Reference mAs = 130 mAs	NI = 20/26/32 HU	NI = 20/26/32 HU
Tube charge (mAs)[2]	100/142/215[3]	88/125/182[3]	110/164/225	96/124/160
CTDI$_{vol}$ (mGy)[2]	6.8/9.6/14.6	5.9/8.5/12.2	7.8/11.6/16	6.7/8.6/11.1
Image reconstruction parameters				
Section thickness/interval	2/1 mm	2/1 mm	2.5/1.25 mm	2.5/1.25 mm
			0.625/0.625 mm[4]	
Pixel size (mm)[2]	0.625/0.723/0.820	0.625/0.723/0.820	0.625/0.625/0.820	0.625/0.723/0.820
Display FOV (mm)[2]	320/370/420	320/370/420	320/370/420	320/370/420
Kernel	B30f (FBP)/I30f (IR)	B30f (FBP)/I30f (IR)	Standard	Standard
Image reconstruction algorithms				
FBP	FBP	FBP	FBP	n.a.[5]
Statistical IR	SAFIRE 3	n.a.	ASiR 50%	n.a.
Partial model-based IR	n.a.	ADMIRE 3	n.a.	ASiR-V 50%
Full model-based IR	n.a.	n.a.	Veo	n.a.

[1] ATCM settings correspond to quality reference mAs for Siemens systems and noise index (NI) values for GE systems. NI values correspond to the small, medium, and large phantom, respectively.

[2] Tube charge, CTDI$_{vol}$, pixel size, and display FOV all varied with the phantom size. Values correspond to the small, medium, and large phantom, respectively.

[3] The mAs values for the two Siemens systems are effective mAs (mAs/pitch).

[4] For the full model-based IR (Veo), 0.625/0.625 mm was the only section thickness/interval reconstruction available at the CT console at the time of the study.

[5] Not available, ASiR-V at a level of 0% was used instead.

From each dataset, because of the different display FOV used depending on the phantom diameter, we extracted square regions of interest (ROIs) containing the low-contrast lesion, which measured 22×22, 18×18, and 16×16 pixels for the small, medium, and large phantoms, respectively. We extracted a total of 60 ROIs per sphere diameter and for the various acquisition/reconstruction conditions. We additionally extracted 300 ROIs of noise only from images of the homogeneous module for each acquisition/reconstruction condition. Due to the non-stationarity of IR algorithms, we selected ROIs of noise only at the same position than signal-containing ROIs.

To provide an objective quantitative metric, we used an anthropomorphic mathematical model observer designed to assess the performance of a classification task (signal present/signal absent). This method is based on a statistical decision theory that aims to mimic the performance of human observers in the detection of low-contrast structures within an image [15–17]. We chose to use the channelized Hotelling observer (CHO) with 10 dense differences of Gaussian channels as a mathematical model observer [18]. The radial frequency of the jth channel is given by .Eq. (1):

$$U_j(\rho) = \exp\left[-\frac{1}{2}\left(\frac{\rho}{W\sigma_j}\right)^2\right] - \exp\left[-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2\right] \tag{1}$$

In this equation, ρ is the spatial frequency, the multiplicative factor W is the channel bandwidth, and σ_j is the standard deviation of the jth channel with σ_j defined from a starting value σ_0 by $\sigma_j = \sigma_0 \alpha^j$. In this study, W = 1.67, $\sigma_0 = 0.005$, and $\alpha = 1.4$ [19].

Each vectorized ROI g, containing either a signal or noise only, was channelized using Eq. (2):

$$\mathbf{v} = \mathbf{U}^T \mathbf{g} \tag{2}$$

where U is a matrix with columns corresponding to channel vectors.

As a human observer, the mathematical observer attributes a grade (also called decision variable λ) to each channelized image. Eq. (3) was used to compute λ:

$$\lambda = \mathbf{w}^T \mathbf{v} \tag{3}$$

And the template w was computed using Eq. (4):

$$\mathbf{w} = \mathbf{K_V^{-1}}(< \mathbf{v_s} > - < \mathbf{v_n} >) \tag{4}$$

In this equation, K_v is the covariance matrix of the 300 channelized signal-absent ROIs, $< v_s >$ is the mean of the 80 channelized ROIs with signal, and $< v_n >$ is the mean of the 300 channelized ROIs with noise only. Using this method, the model observer was trained to calculate the template, and was tested for calculating the decision variables using the same images, which could overestimate the performances (positive bias). The number of signal-absent ROIs and signal-present ROIs were chosen to avoid singularity issues when dealing with the estimation of the covariance matrix inversion and to minimize bias due to the use of a limited number of images [20].

Higher decision variables corresponded to higher confidence of the model observer regarding the presence of a low-contrast lesion in the image. All the decision variables were combined in a histogram, as illustrated as an example in Fig. 2a. Varying a threshold of the decision variable enabled the plotting of a receiver operating characteristic (ROC) curve (Fig. 2b), which was the complete outcome of the model observer. The ROC curve was plotted using 200 thresholds. Next, the image quality was assessed by computing the area under the ROC curve (AUC) using a trapezoidal method.

2.4. Statistical analysis

To each dataset, we applied a bootstrapping technique (random selection with replacement) to estimate the reference range of the results of the model observer [21]. We performed a total of 1000 different bootstrapped samplings using 60 signal images and 300 noise-only images. For each randomization, we calculated the ROC curves and the corresponding AUC values. To assess the statistical differences between the two distributions corresponding to FBP and IR outcomes, we calculated the percentage (PC) of IR outcomes superior to FBP, without making any assumption about the underlying distribution of the data. The improvement of LCD for the IR algorithm compared to FBP on the same CT system was considered statistically significant if PC was > 95% (0.95). This level was set arbitrarily, yet in analogy with the significance levels commonly used for p-values, to be even more specific in terms of discrimination, considering that AUCs > 0.90 are
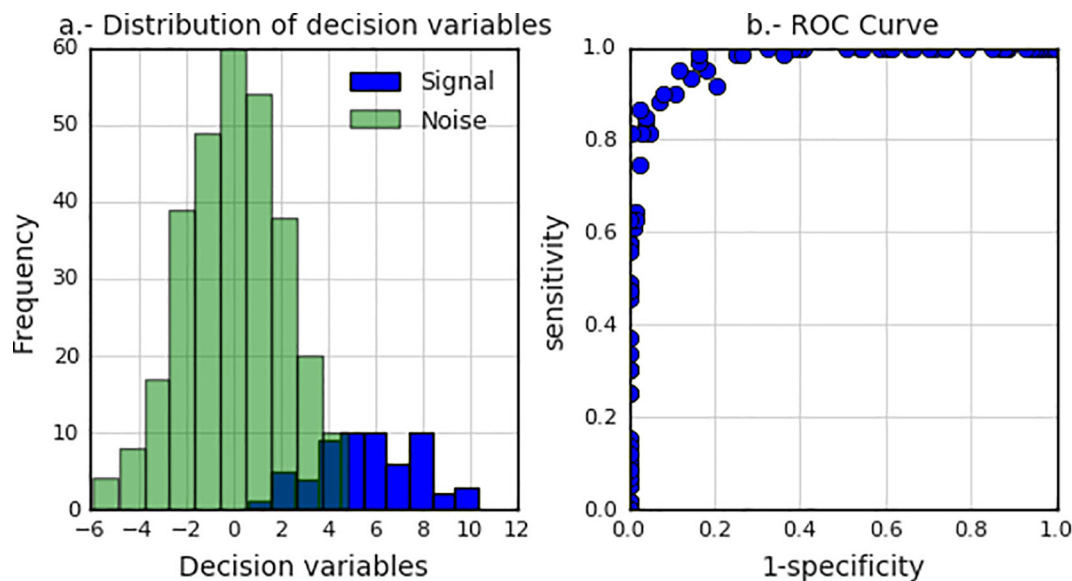
Fig. 2. a) Example of the distribution of decision variables obtained using the CHO model observer for the two categories of images: images with signal (blue bars) and images with noise only (green bars). b) Receiver operating characteristic (ROC) curve showing the outcome of decision variable threshold variation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

rated as excellent [22]. In such cases, we calculated the potential mean improvement.

3. Results

3.1. Low-contrast detectability using FBP

Fig. 3 presents the AUC for each sphere size (5, 6, and 8 mm) obtained using the standard FBP algorithm (except ASiR-V 0% for the Revolution system) as a function of the phantom size. The vertical error bars on each data point represent the 95% confidence intervals. As expected based on the NI settings or ATCM behavior, the AUC decreased with increasing phantom size at each sphere diameter. Moreover, this decreasing performance was systematically higher for spheres of smaller diameter, i.e., with increasing task difficulty. For a sphere diameter of 5 mm, AUC decreased by 18%, 20%, 25%, and 19% for the Flash, Edge, Discovery CT750 HD, and Revolution systems, respectively.

3.2. Impact of the various IR algorithms on LCD performances as a function of phantom size

Fig. 4 presents a pairwise comparison of the results of the CHO model observer for detection of the smallest sphere (5 mm in diameter) using the conventional FBP and available IR algorithms for the three phantom sizes. Table 2 summarizes the PC of IR outcomes that were superior to FBP. For the small phantom, the AUC values were nearly 1 for each CT system. In this case, the detection task was very easy, and LCD was not significantly improved with any of the tested IR algorithms relative to FBP. For the medium phantom, no significant LCD improvement was observed with ASiR 50%, SAFIRE 3, ASiR-V 50%, or ADMIRE 3 compared to FBP. However, with the Discovery CT750 HD system, Veo was associated with a significant AUC improvement of 5.8% compared to FBP. Similarly, with the large phantom, only Veo yielded a significant AUC improvement of 8.5% relative to FBP. When quantitatively comparing all CT systems using a standard FBP algorithm and for each phantom size, we found that the AUC values were nearly equal (differences of ± 0.05), indicating comparable objective image quality despite the dose variations.

3.3. Impact of the various IR algorithms on LCD performances as a function of lesion size

Fig. 5 presents the results of the CHO model observer as a function of sphere size with the various IR algorithms. These results were only obtained for the large phantom, as this was the most discriminating case. As expected, the AUC values systematically increased with increasing sphere diameter. Table 3 presents a pairwise comparison of the IR and FBP algorithms for each tested CT system and each sphere diameter. With the smallest sphere diameter (5 mm), LCD was not significantly improved with the tested IR algorithms compared to FBP, except with Veo. With the 6-mm sphere, slight LCD improvements were noted for ADMIRE 3 (2% mean improvement) and ASiR 50% (2.2% mean improvement). With the 8-mm sphere, we found that LCD was slightly improved for SAFIRE 3 (0.9%), ADMIRE 3 (1.1%), and ASiR-V 50% (1.6%). Surprisingly, ASiR 50% showed no benefit with the 8-mm spheres. The greatest benefit was observed with the full model-based IR algorithm Veo, which was available on the Discovery CT750 HD system, with LCD increased by 8.5% with the 5-mm spheres and by 5.2% for the 6-mm spheres, despite the fourfold thinner slices. Notably, the Veo algorithm achieved a higher relative LCD gain with the more difficult tasks (i.e., smaller sphere diameter).

4. Discussion

In our study, we used a task-based quantitative image quality assessment approach to investigate how various generations of IR algorithms impacted LCD with different simulated patient effective abdominal diameters. We found that with each tested CT system, LCD decreased with increasing phantom size. This was expected since, in the dose optimization process proposed by Siemens, a reduction of LCD in larger patients is considered acceptable to avoid high-dose exposure. For GE systems, technical charts have been proposed to adapt NI values to the individual patient body habitus to avoid non-diagnostic image quality for thinner patients and overexposure for larger patients [5]. Moreover, the technical limitations of X-ray tube power could also limit the feasibility of equivalent noise levels for all patient morphotypes. Using these methodologies, it appears that although dose increased when the phantom's diameter increased, the LCD did not remain constant with all phantom sizes. This implies that diagnostic requirements
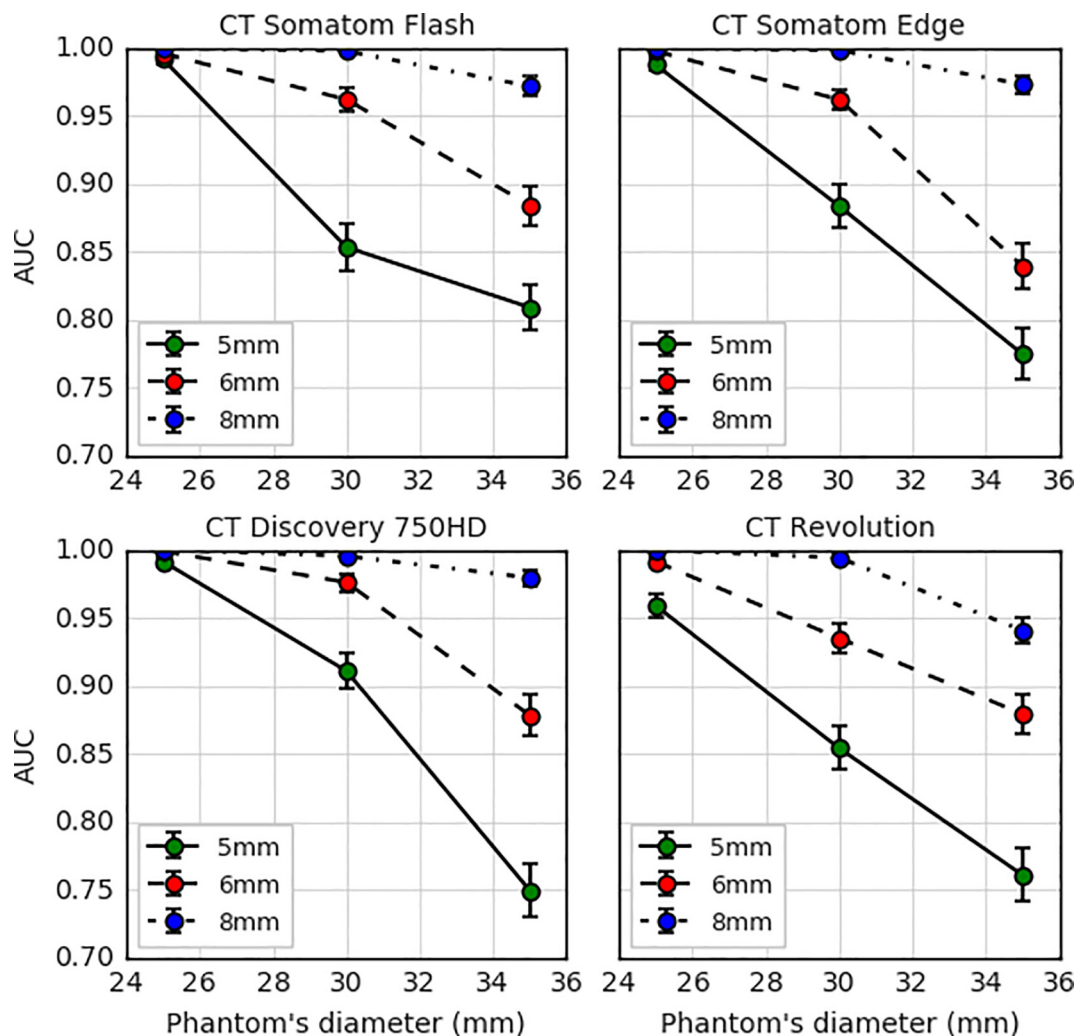
Fig. 3. Low-contrast detectability assessed by means of the area under the receiver operating characteristic curve (AUC) for each computed tomography (CT) system for the three sphere diameters as a function of the phantom diameter for filtered back-projection images.
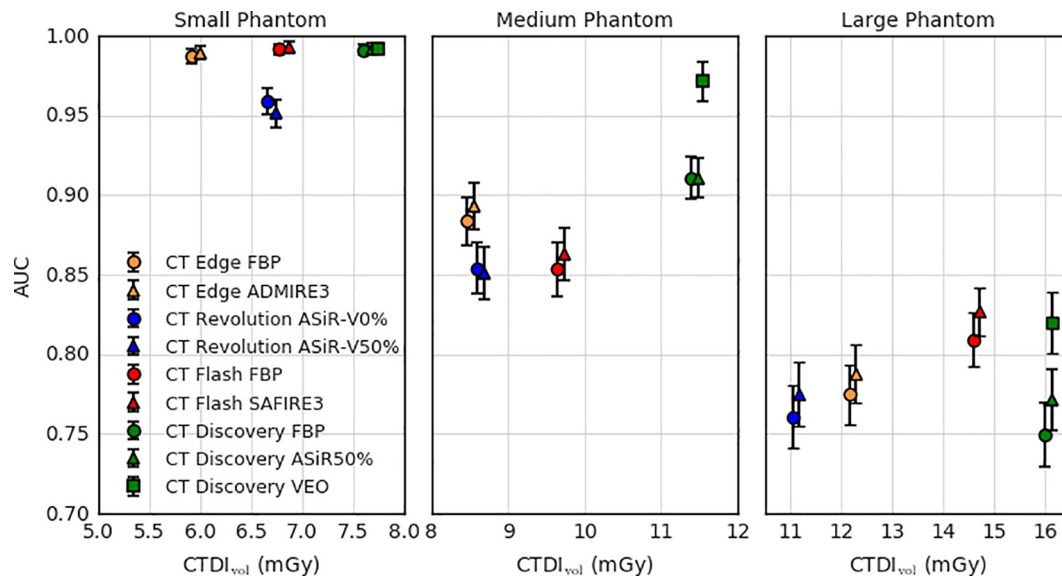

Fig. 4. Impact of the image reconstruction algorithms on low-contrast detectability for 5-mm-diameter spheres and the three phantom sizes. Note that each CT system is reported as a function of the $CTDI_{vol}$ routinely applied in clinical practice.

Table 2

Percentage of iterative reconstruction (IR) outcomes that were superior to filtered back-projection (FBP) for the three phantom sizes for the smallest spheres (5-mm diameter).

Phantom size	Small	Medium	Large
CT Edge (FBP/ADMIRE 3)	76.7%	81.3%	85.0%
CT Flash (FBP/SAFIRE 3)	74.1%	78.4%	93.9%
CT Revolution (ASiR-V 0%/ASiR-V 50%)	9.9%	40.1%	84.6%
CT Discovery CT750 HD (FBP/ASiR 50%)	65.8%	50.1%	94.3%
CT Discovery CT750 HD (FBP/Veo)	64%	100%*	100%*

* Statistically significant improvement between FBP and IR algorithm.

Table 3

Percentage of iterative reconstruction (IR) outcomes that were superior to filtered back-projection (FBP) for the three lesion sizes for the large phantom.

Lesion size	5 mm	6 mm	8 mm
CT Edge (FBP/ADMIRE 3)	85.0%	95.8%*	99.7%*
CT Flash (FBP/SAFIRE 3)	93.9%	92.5%	97.2%*
CT Revolution (ASiR-V 0%/ASiR-V 50%)	84.6%	94.3%	99.4%*
CT Discovery CT750 HD (FBP/ASiR 50%)	94.3%	99.4%*	85.3%
CT Discovery CT750 HD (FBP/Veo)	100%*	100%*	96.4%*

* Statistically significant improvement between FBP and IR algorithm.

should be defined individually based on both the patient-specific body habitus and clinical task. Thus, an optimization process should account for the patient's body habitus and use an objective image quality metric linked to the clinical task.

Our results further demonstrated that the first and second generations of IR techniques (statistical IR and partial model-based IR, respectively) had no substantial impact on LCD for each phantom size. For the small phantom, the examined task was trivial, and thus no difference was observed between the various image reconstruction algorithms. Notably, the dose levels used enabled a high level of performance (Fig. 4). The use of a lower dose range may have enabled the detection of differences between reconstruction algorithms. For the medium and large phantoms, we still observed no substantial improvement for 5-mm diameter lesions with ASiR 50%, SAFIRE 3, ASiR-

V 50%, or ADMIRE 3 compared to FBP. However, the effectiveness of IR algorithms increased with increasing phantom size. Only the full model-based IR Veo, which includes the physical and optical aspects of the CT system in the image reconstruction process, had a significant impact on LCD for the medium and large phantoms. Since Veo further provides a reduced section thickness in comparison with the other algorithms, we secondarily investigated the effect of the reduced section thickness on LCD by averaging thin Veo slices. For the smallest lesion size, no significant impact was noted for each phantom size. As a matter of fact, partial volume effect that may impair LCD should not be significant when dealing with the detection of a 5 mm sphere using a section thickness of 2.5 mm with a reconstruction interval of 1.25 mm. Nevertheless, this reduced section thickness may further improve the longitudinal spatial resolution and the LCD in the other two
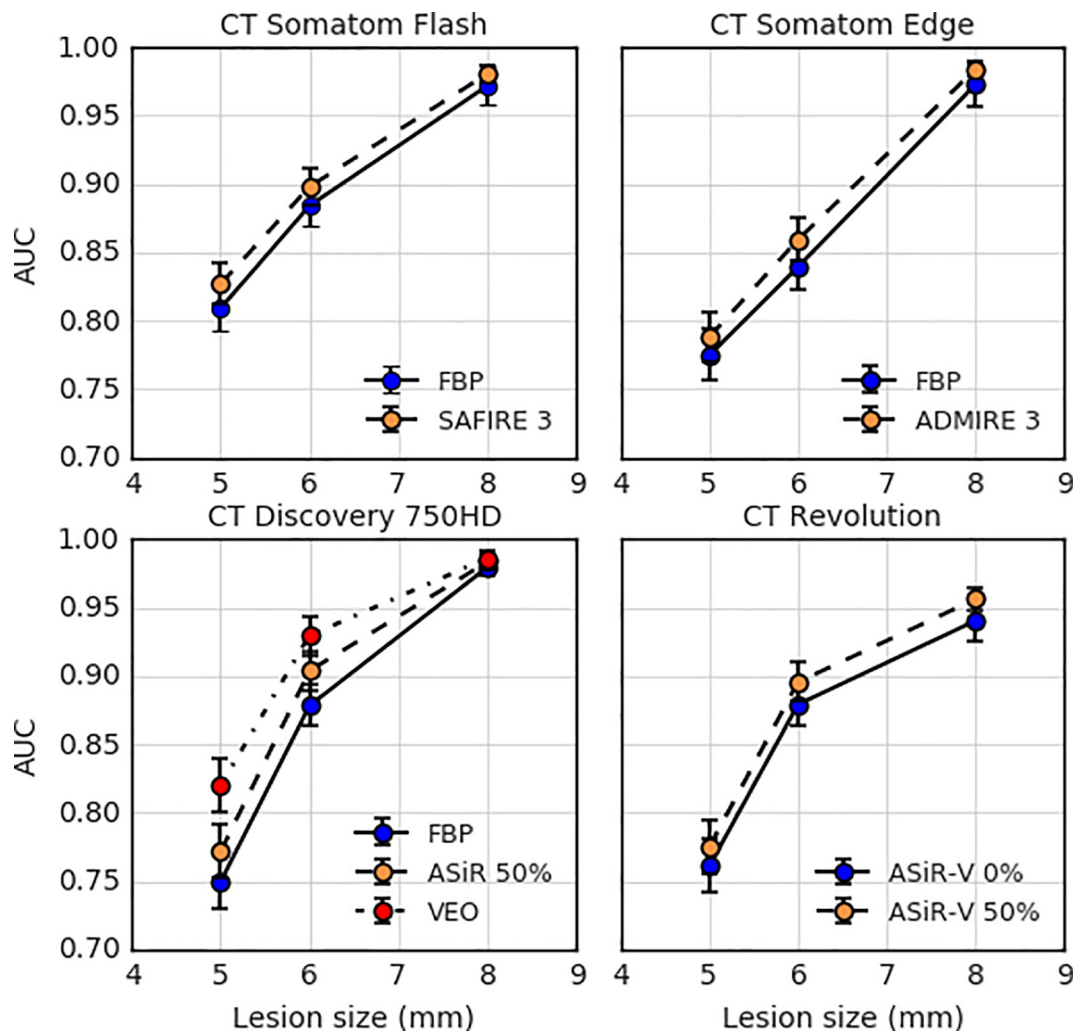

Fig. 5. Low-contrast detectability as a function of lesion size with various image reconstruction algorithms for the large phantom.

reconstruction planes (i.e. coronal and sagittal) when using multiplanar reformations. However, this point was beyond the scope of the present investigation. The main limitation of Veo is that the reconstruction time takes up to 45 min for a single acquisition phase. In summary, our present results clearly demonstrated that, for the selected clinical task, the improvement in LCD by using statistical IRs or partial model-based IRs rather than FBP is very limited. Hence, the potential to achieve substantial dose reduction while maintaining LCD appears to be limited. The use of full model-based IR rather than FBP had greater potential to improve LCD, particularly for large body sizes.

Other investigators have also used CT phantoms to assess the potential improvement of LCD with various IR algorithms. Subjective analyses with human readers (ROC, 4AFC experiment) have indicated that IR techniques (SAFIRE, ASiR, and AIDR) do not statistically improve LCD [9,23–25] or only slightly improve LCD at the same dose level with various IR algorithms [26–28]. These researchers have also applied quantitative image quality parameters (noise and CNR) to validate their use. They report that image noise is significantly reduced with IR algorithms, leading to major CNR improvements. On the other hand, the CNR is not correlated with LCD [29], and should no longer be used as a surrogate of LCD when the noise texture changes, which typically occurs with IR algorithms. This is the case when comparing different CT systems, or when comparing various image reconstruction algorithms on the same CT scanner. Our results confirmed the findings of previous human subjective analyses of statistical IR (ASiR and SAFIRE) [26,27] and partial model-based IR (ASiR-V and ADMIRE) [25,28] using an objective methodology based on a task-specific concept.

Other studies have also applied task-based image quality metrics to LCD assessment using Fourier transform metrics (noise power spectrum and modulation transfer function) to compute a detectability index [30–32] or by directly computing mathematical model observers [16,28,33]. To compute Fourier transform metrics, imaging systems must be linear and shift invariant [34]. However, this is no longer the case with IR algorithms; thus, this method must be applied with caution in such contexts. In our study, we applied a mathematical model observer directly in the image space domain. Our results confirmed the findings of a previous study using Fourier transform metrics for ASiR and Veo [31]. ASiR did not improve LCD as compared to FBP, whereas Veo allowed better improvement compared to both FBP and ASiR.

Our study has several limitations. First, the anthropomorphic abdominal phantom comprised soft-tissue-equivalent rings that do not accurately simulate large patients with fat around their organs in addition to subcutaneous fat. Rather, the large phantom simulates a patient with a higher X-ray attenuation, and we presume that LCD improvement with IR algorithms would be identical or less significant using a phantom with fat-equivalent rings. Second, the technological differences between CT systems and the different ATCM settings, which are based on different principles between the two manufacturers, led to the clinical use of various CTDI$_{vol}$. This was not a problem since the goal of our study was not to benchmark CT systems but rather to investigate the impact of IR over FBP using routine clinical protocols and settings. Despite these differences, we noted that noise levels and LCD of the FBP acquisitions for a common and clinically relevant task were comparable for both manufacturers. Third, we did not investigate the impact of the strength levels of the various IR algorithms on LCD; we only selected one of the most commonly used strength levels proposed by each manufacturer for abdominal CT protocols. Finally, the chosen task (detection of a low-contrast structure at a known position in a homogeneous background) is very simple compared to the clinical reality. However, even in this simple condition, the benefit of IR appears marginal and we should not expect major improvements when dealing with textured backgrounds [35]. To go a step further, we may, in the future, add complexity to the task, and assess the performance of IR algorithms to correctly locate a signal or estimate the size and shape of a lesion [36,37].

5. Conclusions

In conclusion, in routinely used abdominal CT protocols, switching from FBP to IR algorithms does not significantly improve LCD. Dose optimization should be performed by first defining the diagnostic requirement for a relevant clinical task and the patient's specific body habitus, and by subsequently evaluating image quality criteria linked to this task before drastically reducing dose levels.

Source of funding

Conflict of interest

The authors have no relevant conflict of interest to disclose.

References

[1] Schauer DA, Linton OW. National council on radiation protection and measurements report shows substantial medical exposure increase. Radiology 2009;253:293–6. http://dx.doi.org/10.1148/radiol.2532090494.

[2] Lee CH, Goo JM, Ye HJ, Ye S-J, Park CM, Chun EJ, et al. Radiation dose modulation techniques in the multidetector CT era: from basics to practice. RadioGraphics 2008;28:1451–9. http://dx.doi.org/10.1148/rg.285075075.

[3] Söderberg M, Gunnarsson M. Automatic exposure control in computed tomography – an evaluation of systems from different manufacturers. Acta Radiol 2010;51:625–34. http://dx.doi.org/10.3109/02841851003698206.

[4] McCollough CH. Automatic exposure control in ct: are we done yet? Radiology 2005;237:755–6. http://dx.doi.org/10.1148/radiol.2373051151.

[5] Favazza CP, Yu L, Leng S, Kofler JM, McCollough CH. Automatic exposure control systems designed to maintain constant image noise: effects on computed tomography dose and noise relative to clinically accepted technique charts. J Comput Assist Tomogr 2015;1. http://dx.doi.org/10.1097/RCT.0000000000000221.

[6] Patino M, Fuentes JM, Singh S, Hahn PF, Sahani DV. Iterative Reconstruction Techniques in Abdominopelvic CT: technical Concepts and Clinical Implementation. Am J Roentgenol 2015;205:W19–31. http://dx.doi.org/10.2214/AJR.14.13402.

[7] Trattner S, Pearson GDN, Chin C, Cody DD, Gupta R, Hess CP, et al. Standardization and Optimization of CT Protocols to Achieve Low Dose. J Am Coll Radiol 2014;11:271–8. http://dx.doi.org/10.1016/j.jacr.2013.10.016.

[8] Verdun FR, Racine D, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, Veldkamp WJH, et al. Image quality in CT: from physical measurements to model observers. Phys Medica Eur J Med Phys 2015;31:823–43. http://dx.doi.org/10.1016/j.ejmp.2015.08.007.

[9] Goenka AH, Herts BR, Dong F, Obuchowski NA, Primak AN, Karim W, et al. Image noise, CNR, and detectability of low-contrast, low-attenuation liver lesions in a phantom: effects of radiation exposure, phantom size, integrated circuit detector, and iterative reconstruction. Radiology 2016;280:475–82.

[10] Schindera ST, Odedra D, Raza SA, Kim TK, Jang H-J, Szucs-Farkas Z, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? Radiology 2013;269:511–8.

[11] Greffier J, Pereira F, Macri F, Beregi J-P, Larbi A. CT dose reduction using Automatic Exposure Control and iterative reconstruction: a chest paediatric phantoms study. Phys Med 2016;32:582–9. http://dx.doi.org/10.1016/j.ejmp.2016.03.007.

[12] Racine D, Viry A, Becce F, Schmidt S, Ba A, Bochud FO, et al. Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners. Med Phys 2017;44:e153–63. http://dx.doi.org/10.1002/mp.12224.

[13] Viry A, Racine D, Ba A, Becce F, Bochud FO, Verdun FR. Characterization of a CT unit for the detection of low contrast structures. In: Kupinski MA, Nishikawa RM, editors. 2017. p. 101361C. https://doi.org/10.1117/12.2250529.

[14] Marrero JA, Rajender Reddy K. ACG clinical guideline: the diagnosis and management of focal liver lesions. Am J Gastroenterol 2014;109:1328–47. http://dx.doi.org/10.1038/ajg.2014.213.

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

[16] Ott JG, Ba A, Racine D, Viry A, Bochud FO, Verdun FR. Assessment of low contrast detection in CT using model observers: developing a clinically-relevant tool for characterising adaptive statistical and model-based iterative reconstruction. Z Für Med Phys 2016. http://dx.doi.org/10.1016/j.zemedi.2016.04.002.

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

[18] Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci U S A 1993;90:9758–65.

[19] Abbey CK. Human-and model-observer performance in ramp-spectrum noise:

effects of regularization and object variability. JOSA A 2001;18:473–88.

[20] 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;34:453–64. http://dx.doi.org/10.1109/TMI.2014.2360496.

[21] Gagne RM, Gallas BD, Myers KJ. Toward objective and quantitative evaluation of imaging systems using images of phantoms. Med Phys 2006;33:83. http://dx.doi.org/10.1118/1.2140117.

[22] Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. J Pediatr Psychol 2014;39:204–21. http://dx.doi.org/10.1093/jpepsy/jst062.

[23] Miéville FA, Gudinchet F, Brunelle F, Bochud FO, Verdun FR. Iterative reconstruction methods in two different MDCT scanners: physical metrics and 4-alternative forced-choice detectability experiments – A phantom approach. Phys Med 2013;29:99–110. http://dx.doi.org/10.1016/j.ejmp.2011.12.004.

[24] Schindera ST, Odedra D, Raza SA, Kim TK, Jang H-J, Szucs-Farkas Z, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? Radiology 2013;269:511–8. http://dx.doi.org/10.1148/radiol.13122349.

[25] Euler A, Stieltjes B, Szucs-Farkas Z, Eichenberger R, Reisinger C, Hirschmann A, et al. Impact of model-based iterative reconstruction on low-contrast lesion detection and image quality in abdominal CT: a 12-reader-based comparative phantom study with filtered back projection at different tube voltages. Eur Radiol 2017;27:5252–9. http://dx.doi.org/10.1007/s00330-017-4825-9.

[26] Goenka AH, Herts BR, Obuchowski NA, Primak AN, Dong F, Karim W, et al. Effect of reduced radiation exposure and iterative reconstruction on detection of low-contrast low-attenuation lesions in an anthropomorphic liver phantom: an 18-reader study. Radiology 2014;272:154–63. http://dx.doi.org/10.1148/radiol.14131928.

[27] McCollough CH, Yu L, Kofler JM, Leng S, Zhang Y, Li Z, et al. Degradation of CT Low-Contrast Spatial Resolution Due to the Use of Iterative Reconstruction and Reduced Dose Levels. Radiology 2015;276:499–506. http://dx.doi.org/10.1148/radiol.15142047.

[28] 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;169:73–7. http://dx.doi.org/10.1093/rpd/ncw020.

[29] Urikura A, Hara T, Ichikawa K, Nishimaru E, Hoshino T, Yoshida T, et al. Objective assessment of low-contrast computed tomography images with iterative reconstruction. Phys Med 2016;32:992–8. http://dx.doi.org/10.1016/j.ejmp.2016.07.003.

[30] 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: towards task-based assessment of CT performance. Med Phys 2012;39:4115–22. http://dx.doi.org/10.1118/1.4725171.

[31] Samei E, Richard S. Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology: CT task-based performance metrology. Med Phys 2014;42:314–23. http://dx.doi.org/10.1118/1.4903899.

[32] 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;59:4047–64. http://dx.doi.org/10.1088/0031-9155/59/4/4047.

[33] 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. http://dx.doi.org/10.1016/j.ejmp.2015.09.011.

[34] 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. http://dx.doi.org/10.1088/0031-9155/60/2/R1.

[35] Solomon J, Ba A, Bochud F, Samei E. Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms: voxel-based 3D printed textured phantoms. Med Phys 2016;43:6497–506. http://dx.doi.org/10.1118/1.4967478.

[36] Solomon J, Mileto A, Nelson RC, Roy Choudhury K, Samei E. Quantitative features of liver lesions, lung nodules, and renal stones at multi-detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology 2016;279:185–94. http://dx.doi.org/10.1148/radiol.2015150892.

[37] Takata T, Ichikawa K, Mitsui W, Hayashi H, Minehiro K, Sakuta K, et al. Object shape dependency of in-plane resolution for iterative reconstruction of computed tomography. Phys Med 2017;33:146–51. http://dx.doi.org/10.1016/j.ejmp.2017.01.001.