

Technology Characterization Through Diverse Evaluation Methodologies: Application to Thoracic Imaging in Photon-Counting Computed Tomography

Jayasai R. Rajagopal, PhD,*† Fides R. Schwartz, MD,‡ Cindy McCabe, BS,* Faraz Farhadi, BS,‡§
Mojtaba Zarei, MS,* Francesco Ria, DMP,* Ehsan Abadi, PhD,* Paul Segars, PhD,*
Juan Carlos Ramirez-Giraldo, PhD,|| Elizabeth C. Jones, MD,† Travis Henry, MD,‡
Daniele Marin, MD,‡ and Ehsan Samei, PhD*

Objective: Different methods can be used to condition imaging systems for clinical use. The purpose of this study was to assess how these methods complement one another in evaluating a system for clinical integration of an emerging technology, photon-counting computed tomography (PCCT), for thoracic imaging.

Methods: Four methods were used to assess a clinical PCCT system (NAEOTOM Alpha; Siemens Healthineers, Forchheim, Germany) across 3 reconstruction kernels (Br40f, Br48f, and Br56f). First, a phantom evaluation was performed using a computed tomography quality control phantom to characterize noise magnitude, spatial resolution, and detectability. Second, clinical images acquired using conventional and PCCT systems were used for a multi-institutional reader study where readers from 2 institutions were asked to rank their preference of images. Third, the clinical images were assessed in terms of in vivo image quality characterization of global noise index and detectability. Fourth, a virtual imaging trial was conducted using a validated simulation platform (DukeSim) that models PCCT and a virtual patient model (XCAT) with embedded lung lesions imaged under differing conditions of respiratory phase and positional displacement. Using known ground truth of the patient model, images were evaluated for quantitative biomarkers of lung intensity histograms and lesion morphology metrics.

Results: For the physical phantom study, the Br56f kernel was shown to have the highest resolution despite having the highest noise and lowest detectability. Readers across both institutions preferred the Br56f kernel (71% first rank) with a high interclass correlation (0.990). In vivo assessments found superior detectability for PCCT compared with conventional computed tomography but higher noise and reduced detectability with increased kernel sharpness. For the virtual imaging trial, Br40f was shown to have the best performance for histogram measures, whereas Br56f was shown to have the most precise and accurate morphology metrics.

Conclusion: The 4 evaluation methods each have their strengths and limitations and bring complementary insight to the evaluation of PCCT. Although no method offers a complete answer, concordant findings between methods offer affirmatory confidence in a decision, whereas discordant ones offer insight for added perspective. Aggregating our findings, we concluded the Br56f kernel best for high-resolution tasks and Br40f for contrast-dependent tasks.

Key Words: protocol design, photon-counting CT, virtual imaging trial, thoracic imaging

(*J Comput Assist Tomogr* 2024;00: 00–00)

Before new imaging technologies can be adapted into clinical use, their best implementation must be determined. Given the high throughput of the clinical workflow, protocols designed for new technologies need to be general so that the workflow is unencumbered. However, protocols must also be tailored to specific clinical goals. Clinical cases may further require modification of a protocol on a case-by-case basis. Thus, the challenge in designing protocols is to provide the best clinical image quality by providing a general guideline for routine evaluation while also enabling flexibility to customize options for specific clinical tasks.

There are several approaches to compare different options for new systems when designing protocols, each of which offers some advantages and disadvantages. Evaluation with physical phantoms provides a direct characterization of system properties.¹ Because phantoms can be repeatedly imaged and have consistent internal structure and geometry, these evaluations can be performed systematically to consider a variety of conditions. However, phantom evaluations are limited in their complexity and clinical realism when compared with the variability that can be expected across a patient population. Clinical evaluations can be performed on patient images through reader studies using expert readers that assess the images objectively or subjectively taking into account clinical needs and reader preferences. However, reader studies are logistically more involved including cost, time, and ethical considerations. Another form of clinical evaluation is in vivo image quality characterization using patient images. Such methods are able to measure image quality with high realism but share the ethical challenges of reader studies with regard to repeated scans. Another method of evaluation is the use of virtual imaging trials (VITs),² which use virtual patient models and scanner models to simulate medical images for targeted conditions. Because these virtual images have a known ground truth and can be acquired repeatedly, they enable assessment of factors that would not be possible with other methods³ but face challenges in the form of computational resources.

One example of a recent technological development currently in the process of clinical adoption is the incorporation of photon-counting detectors within computed tomography (CT)

From the *Center for Virtual Imaging Trials, Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, Durham, NC; †Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD; ‡Duke University Health System, Department of Radiology, Duke University Medical Center, Durham, NC; §Geisel School of Medicine, Dartmouth College, Hanover, NH; and ||Siemens Healthineers USA, Malvern, PA. Received for publication February 2, 2024; accepted February 11, 2024.

Correspondence to: Jayasai R. Rajagopal, PhD, Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, 2424 Erwin Dr, Ste 302, Durham NC, 27705 (e-mail: jayasai.rajagopal@duke.edu). Supported in part from a grant from the National Institutes of Health (P41EB028744).

The content of this article does not necessarily reflect the views or policies of the Department of Health and Human Services, nor do mention of trade names, commercial products, or organizations imply endorsement by the US Government. J.C.R.G. is an employee of Siemens Healthineers. The authors unaffiliated with Siemens had full control over the data and information presented in this article. P.S. is an employee of Siemens Healthineers. E.S. lists relationships with the following entities unrelated to the present publication: GE, Bracco, Imaloxig, 12Sigma, Metis Health Analytics, Cambridge University Press, and Wiley.

Copyright © 2024 Wolters Kluwer Health, Inc. All rights reserved.
DOI: 10.1097/RCT.0000000000001608

systems, replacing conventional energy-integrating detectors (EIDs). Conventional CT is commonly the first-line modality for imaging of numerous thoracic applications due to its high spatial resolution, short examination time, and general availability, demonstrated in the characterization of diffuse lung disease,⁴ lung lesions and masses,⁵ and COVID-19.⁶ Computed tomography has also proven valuable for use in monitoring for lung cancer,⁷ in emergency care,⁸ and in the intensive care unit.^{9–11} Computed tomography based on EIDs uses a scintillation layer and a photodiode to record photon signals that are integrated in the formation of an image signal. However, CT based on photon-counting detectors uses semiconductors to directly record individual photons at the signal formation.^{12,13} Photon-counting CT (PCCT) has been shown to offer several benefits over EID CT including higher spatial resolution,¹⁴ lower noise especially in low-dose conditions,¹⁵ higher contrast,¹⁶ and spectral stability.¹⁷ Early investigations with PCCT have also shown benefits in improving clinical thoracic imaging by reducing dose in routine scans,¹⁸ including lung cancer screening,¹⁹ and improving characterization of lung disease.^{20,21}

The first PCCT system was recently cleared by the US Food and Drug Administration for clinical use,²² and thus the best clinical use of this technology has become an active area of research. In doing so, as noted above, there exists a host of evaluation methods. In this study, we aimed to deploy a suite of methods (phantom measurements, reader studies, in vivo characterization, and VITs) to assess how they may best complement one another. The study focused on the choice of reconstruction kernel for a generic pulmonary imaging task using 4 evaluation techniques. To elucidate the usefulness of these methods in a clinical imaging setting, we chose to focus on how the choice of reconstruction kernel may be made.

MATERIALS AND METHODS

J.C.R.G. is from the industry. Data were collected, controlled, and analyzed by J.R.R., F.R.S., and C.M. who were not affiliated with industry without potential bias.

Clinical Photon-Counting Scanner

This study used a dual-source clinical PCCT system²³ (NAEOTOM Alpha; Siemens Healthineers, Forchheim, Germany) that has a 50-cm field of view along the primary source-detector subsystem of the scanner. Images were acquired at 120 kV using the Quantum Plus (Qplus) mode, which uses a 144 × 0.4-mm collimation, 0.4 × 0.4-mm in-plane pixels, and 2 detector energy thresholds, which were fixed to 20 and 65 keV. All images were reconstructed using an offline reconstruction program (ReconCT v. 15.0.53098.0; Siemens, Erlangen, Germany) with a 0.6-mm slice thickness with 3 reconstruction kernels (Br40f, Br48f, and Br56f). These kernels were chosen to be comparable with the

standard for thoracic CT scans at our institution. The slice thickness was chosen to match the standard of care. Other acquisition and reconstruction parameters, including tube current and reconstruction field of view, were adjusted for the requirements of each evaluation (Table 1). All evaluations were done on images that contain the entire spectral signal (termed T3D).

Physical Phantom Evaluation

For the physical phantom evaluation, the Mercury Phantom 4.0 (Sun Nuclear, Melbourne, Fla) was imaged. The phantom is constructed of polyethylene (Hounsfield unit [HU] value of −90 at 120 kV) at 5 sizes (16-, 21-, 26-, 31-, and 36-cm diameters), which are composed of 2 sections: a uniform 3-cm axial section for noise assessment and a 3-cm axial section containing inserts (2.54 cm in diameter) representing clinically relevant materials (air, bone, fat, iodine, water). Image quality evaluation was done using the air insert for task-specific characterization. The phantom was scanned on the PCCT system in Qplus mode at 152 mAs and reconstructed to a 30-cm field of view with a 1024 × 1024 matrix size. Images were evaluated using a software package, imQuest²⁴ (version 7.0; Duke University, Durham, NC), which enables analysis of target regions of interest (ROIs) according to American Association of Physicists in Medicine Task Group-233.²⁵

Noise power spectrum (NPS) was used as a measure of the magnitude and texture of noise.^{26,27} Twelve square ROIs (32 × 32 pixels, 0.9 × 0.9 cm²) were placed equiangularly along a circle (1.8 cm) within the uniform region of the phantom at each size. The 2-dimensional NPS was then calculated and radially binned and averaged to calculate a 1-dimensional NPS.²⁸ The NPS was further normalized by dividing by noise magnitude to create the final normalized NPS (NNPS). Noise was represented by 2 figures of merit: the noise magnitude (ie, the standard deviation of pixel values) and the peak NPS frequency.

Spatial resolution was measured using the task transfer function (TTF),²⁹ which is analogous to the modulation transfer function but accounts for task-dependent features of resolution including local contrast and background noise. The circular rod method^{30,31} was used to estimate the TTF, where a circular ROI was used to align the center of the rod in the image with an idealized insert. The ensemble edge spread function was then calculated by plotting pixel value against radial angle. After binning to reduce the influence of noise, the derivative of the edge spread function was taken to generate a line spread function, which was Fourier transformed into the final TTF. The TTF f_{50} , the spatial frequency at which the TTF reaches 50% of its maximum value, was taken as a figure of merit for spatial resolution.

Contrast-to-noise ratio (CNR) and detectability index (d') were used as measures of overall image quality. Contrast-to-noise ratio was calculated as absolute value of the ratio of noise

TABLE 1. Summary of Acquisition and Reconstruction Parameters for Each Evaluation Method

Data Set	Physical Phantom Evaluation	Clinical Acquisition Evaluation	Virtual Imaging Evaluation
Acquisition mode (collimation [mm])		Quantum Plus (144 × 0.4)	
Tube voltage, kV		120	
Energy thresholds, keV		20, 65	
Dose	152 mAs	3.0–8.8 mGy	45 mAs
Reconstruction kernel		Br40f, Br48f, Br56f	
Reconstruction field of view, cm × cm	30.0	29.6–36.0	30.0

Dose is reported in terms of mAs for phantom studies and CTDI_{vol} for clinical studies.

magnitude and local contrast derived from the TTF measurement. Detectability index, as a measure of signal detection in the presence of noise,³² was characterized based on a non-prewhitening matched filter observer model with eye filter model.^{31,33} For calculating d' , the task was defined to be the depiction of a 1-mm-diameter image feature with a material-specific contrast level.

Clinical Reader and In Vivo Image Quality Evaluation

The clinical reader evaluation used images taken from a Health Insurance Portability and Accountability Act-compliant prospective study approved by the institutional review board (Pro00107998). All patients consented to having a same-day PCCT scan in addition to their clinical scan on an EID system; no patients were excluded. Original clinical scans were acquired on the following systems: SOMATOM Definition Flash, Definition Force (Siemens Healthineers, Forchheim, Germany), or Revolution Apex (GE Healthcare, Waukesha, Wis). The EID scans were used as a reference for the current clinical standard. Scan protocols are summarized in Table 2.

Nine physicians across 2 institutions (8–24 years of experience interpreting CT) participated as readers in a 4-alternative forced rank study. Photon-counting CT images were reconstructed with the same field of view as each EID scan on a patient-by-patient basis with all 3 kernels. Target slices were extracted with a focus on matching anatomy and clinically relevant pathology and cropped to 350 × 350 pixels. A custom Web application (JavaScript version 1.7) was designed to display a quartet of images with a fixed window width/window level of 1500/–500 to represent a lung window. Readers were blinded to the source of each image and asked to rank the 4 images (EID, PCCT Br40f, PCCT Br48f, and PCCT Br56f) in order of preference with a rank of 1 representing the most preferred and a rank of 4 representing the least preferred images. Statistical analysis was done using MATLAB (v2021a; MathWorks, Natick, Mass). Interrater reliability was calculated using Shrout-Fleiss intraclass correlation coefficient.³⁴ The Wilcoxon rank-sum test was used to compare results across readers and patients to estimate overlap between results.

For the same patient cohort and images included in the clinical reader evaluation, an automatic image quality evaluation was performed. In particular, using previously validated algorithms,^{28,35–37} the average noise magnitude in terms of Global Noise Index (GNI), NNPS, and TTF was calculated. The detectability index (d') was calculated based on the Fisher-Hotelling observer model for a reference 0.8-mm feature at 700 HU. The d' for the 3 PCCT kernels was compared in terms of percentage difference with the EID CT.

VIT Evaluation

Simulated acquisitions were performed using DukeSim (Duke University, Durham, NC),³⁸ a simulator that enables rapid simulation of realistic CT images using computational phantoms.

The simulator can replicate scanner-specific geometry, physics, and protocol settings³⁹ and has recently been extended and validated for simulation of PCCT systems.⁴⁰ For this work, the geometry and protocol settings were matched with the same PCCT scanner used for the experimental and clinical scans. Simulations were done using 45 mAs, and images were reconstructed to a 30.0-cm field of view and a 512 × 512 matrix size. The computational phantom used for the study was taken from the XCAT library.⁴¹ A male phantom, with a body mass index (BMI) of 29, was used as a representative patient. Six lesions (8–9 mm)⁴² were inserted into the virtual patient, 3 in each lung at superior, middle, and inferior positions. The phantom was rendered at 10 different respiratory phases to provide intrapatient variability of lung shapes and volumes during an acquisition. The original phantom was also rendered in 4 positions by displacing it in horizontal and angular directions by ±2 and ±10°cm to represent potential positional variability.

Analysis of simulated images was performed in 2 ways. First, changes in lung density due to change in respiratory phases or positional displacement were evaluated using a histogram analysis. Ground-truth phantoms under each condition were masked to identify lung tissue and exclude all other voxels including lung lesions. Lung parenchyma voxels were segmented using the masks. The absolute error between ground-truth histograms and histograms under each kernel and condition was calculated and represented in terms of mean and standard deviation. Kernel pairs were compared using a 2-way t test with significance for a P value < 0.05.

Second, the images were evaluated for quantifying morphological radiomics features of the inserted lesions. The measurements were done using a previously described method.⁴³ Square ROIs were drawn around each lesion, which was then segmented using an active contour technique without edge smoothing.⁴⁴ The segmented masks were evaluated with the Pyradiomics package⁴⁵ to calculate morphological metrics. Features were measured on both ground-truth phantoms and reconstructed images. For each radiomics feature, bias was measured as the percentage error between the ground truth and measured feature value across the 3 kernels and 6 lesions. The variability in measurements was calculated as the coefficient of variation across the 6 lesions.

RESULTS

Physical Phantom Results

Physical phantoms results showed the Br40f kernel to yield the lowest mean noise magnitude (22.1 HU) when compared with Br48f (37.2 HU) and Br56f (67.4 HU). The Br40f kernel also had the lowest peak NPS frequency (0.22 mm⁻¹) when compared with Br48f (0.34 mm⁻¹) and Br56f (0.48 mm⁻¹). For all three kernels, noise magnitude increased as a function of phantom size, but the shape of the NNPS remained similar (Fig. 1).

TABLE 2. Summary of High-Resolution Chest CT Protocol Parameters for Each Scanner Used in the Reader Study

	Siemens Flash	Siemens Force	GE Revolution Apex	PCCT
Tube voltage, kV			120	
Image type	keV	keV	keV	T3D
Reconstruction parameters	FBP; Br40f	FBP; Br40f	FBP; standard	FBP; Br40f, Br46f, Br56f
Reconstruction reference level (Qref mAs or noise index)	110 mAs	100 mAs	21	115 mAs
Reconstruction slice thickness, mm	0.6	0.6	0.625	0.2

CT indicates computed tomography; FBP, filtered backprojection; PCCT, photon-counting computed tomography.

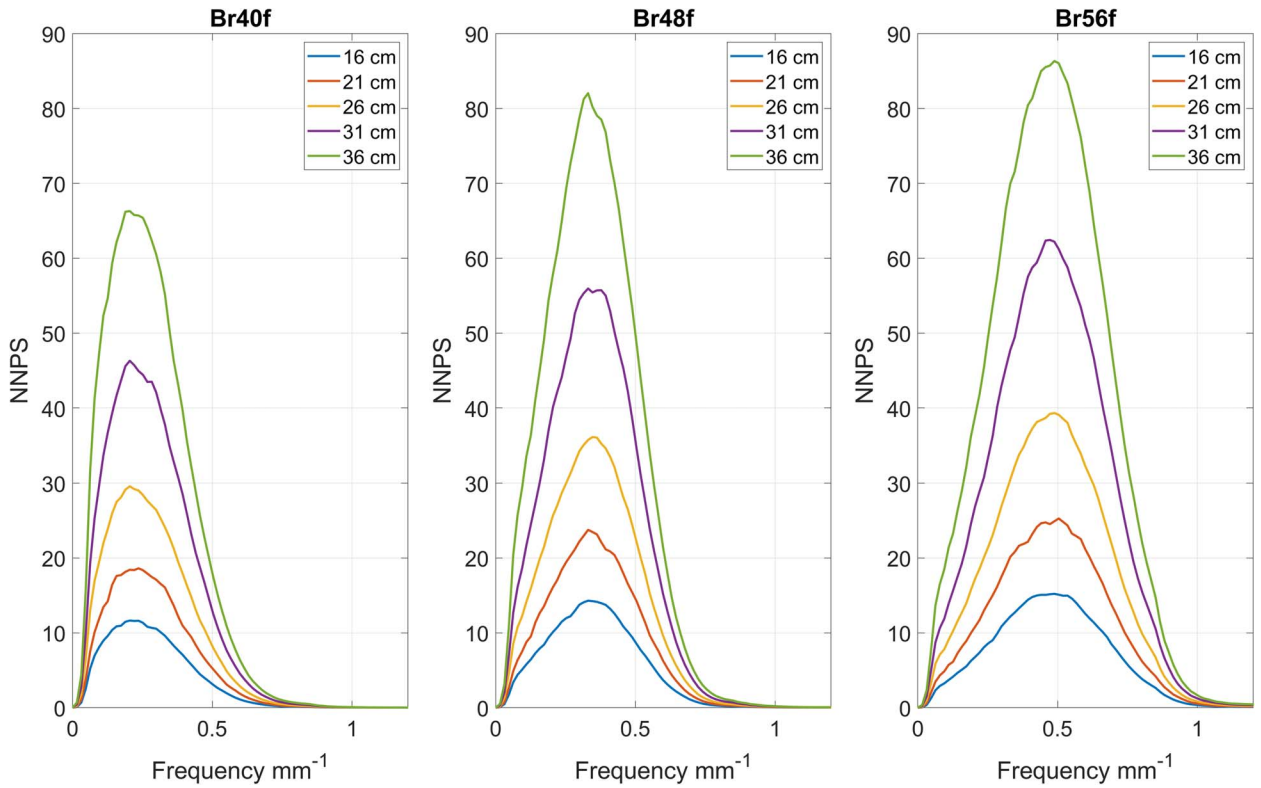


FIGURE 1. Normalized noise power spectra (NNPS) for each kernel (Br40f, left; Br48f, middle; Br56f, right). Phantom sizes separated by color (16 cm, blue; 21 cm, red; 26 cm, yellow; 31 cm, purple; 36 cm, green).

For the TTF measurements (Fig. 2), there was increased variation as the Br56f kernel showed a higher peak than either Br48f or Br40f. In terms of TTF f_{50} , Br40f had a lower value (0.39 mm^{-1}) than Br48f (0.53 mm^{-1}) or Br56f (0.73 mm^{-1}); as expected, the Br56f kernel had a higher spatial resolution at the cost of increased noise across all conditions.

The CNR and detectability index results further reflected these findings. Br40f had the highest mean CNR (59.5) and detectability index (12.0) when compared with Br48f (CNR, 34.9;

d' , 11.3) and Br56f (CNR, 19.1; d' , 10.4). Both CNR and detectability index decreased as a function of size across all 3 kernels (Table 3).

Clinical Observer Study and In Vivo Image Quality Results

The study population consisted of 20 patients (mean age, 70 years; age range, 57–83 years; mean BMI, 28.6; BMI range,

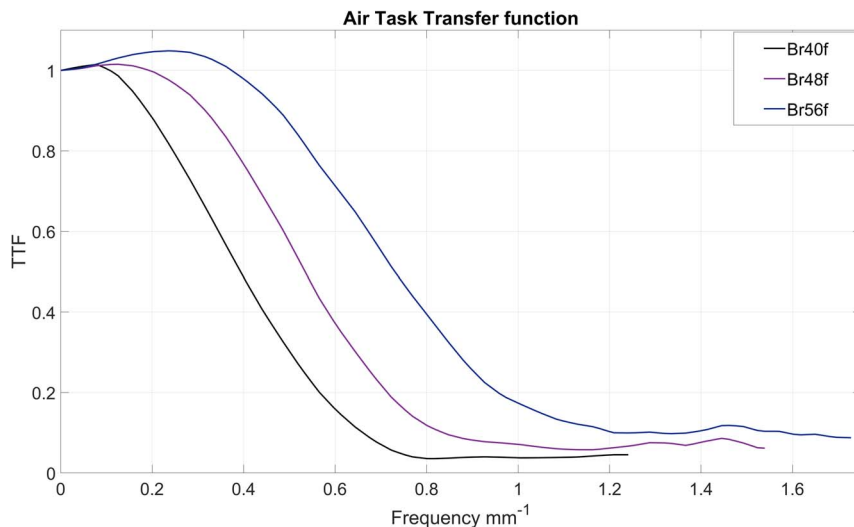


FIGURE 2. Task transfer function (TTF) of the air insert in the Mercury Phantom. Different kernels indicated by color (Br40f, black; Br48f, purple; Br56f, blue). TTF calculated at a dose of 10.9 mGy, for an insert of 3-cm diameter, and at a pixel size of $0.59 \times 0.59 \text{ mm}$.

Downloaded from http://journals.lww.com/jcat by BhDM5fPQKvz2EOumT1QIN4a+KkLHEZQpsIH04XMM0h0CwCk1AW nYQp/IOH-D3I3D00dRy/ITV5FA/C3V/C1y0abggQZxdmKZB7ws= on 04/17/2024

TABLE 3. Contrast-to-Noise Ratio and Detectability Index of the Air Insert of the Mercury Phantom for Each Kernel

Phantom Size, cm	Br40f		Br48f		Br56f	
	CNR	<i>d'</i>	CNR	<i>d'</i>	CNR	<i>d'</i>
16	123.0	25.0	72.3	23.7	39.8	21.9
21	76.3	15.5	44.6	14.7	24.5	13.5
26	47.9	9.6	28.1	9.1	15.4	8.4
31	30.6	6.1	17.9	5.7	9.8	5.3
36	19.7	3.7	11.5	3.4	6.3	3.2

CNR indicates contrast-to-noise ratio; *d'*, detectability index.

19.1–44.4), including 9 men (mean age, 70 years; age range, 59–83 years; mean BMI, 30.9; BMI range, 23.2–39.0) and 11 women (mean age, 70 years; age range, 57–81 years; mean BMI, 26.8; BMI range, 19.1–44.4). Scans on the PCCT system had a lower dose (mean CTDI_{vol}, 5.4 mGy; range, 3.0–8.8 mGy) compared with the EID systems (mean CTDI_{vol}, 7.0 mGy; range,

2.4–11.9). An example of 3 different cases is shown across all 4 imaging conditions (Fig. 3).

Across all readers (Fig. 4), the Br56f kernel was most frequently ranked first, with 71% of all first ranks, whereas EID had the fewest, with 2% of all first ranks. Conversely, EID was most frequently ranked fourth, with 60% of all fourth ranks, and

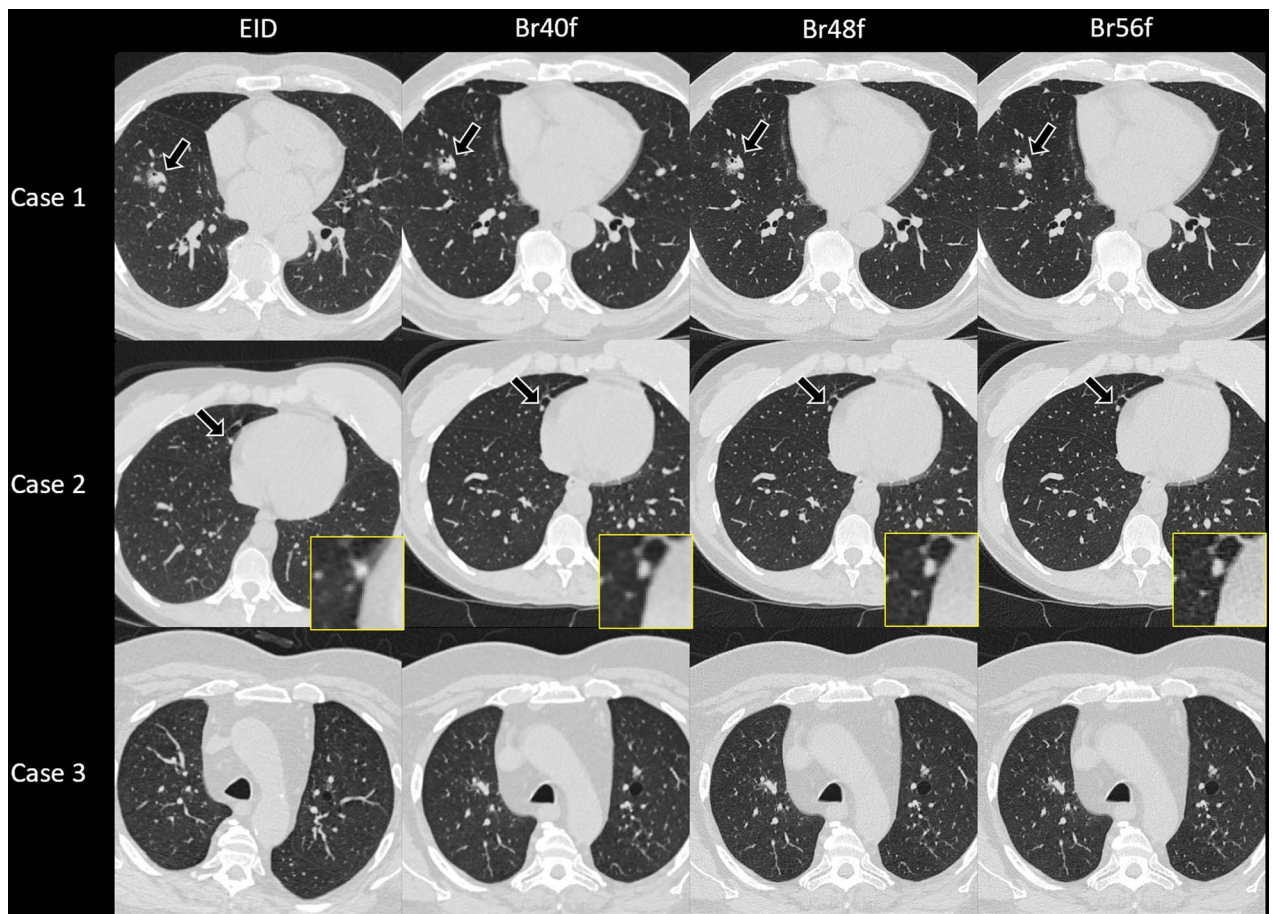


FIGURE 3. Examples of 3 different cases used in the reader study. Case 1 shows a 68-year-old male patient with a body mass index (BMI) of 28 kg/m², with known malignancy in the right lower lobe (black arrow). Case 2 shows images of a 53-year-old woman with a BMI of 22 kg/m² and pulmonary nodules that had been stable for 12 months (black arrow, magnified in yellow box). Case 3 shows images of a 73-year-old man with a BMI of 35 kg/m², who was followed up for small pulmonary nodules in the setting of nonpulmonary malignancy. Columns indicate different image sources (energy-integrating detector, left; Br40f, middle left; Br48f, middle right; Br56f, right). Images presented with a window width/window level of 1500/–500.

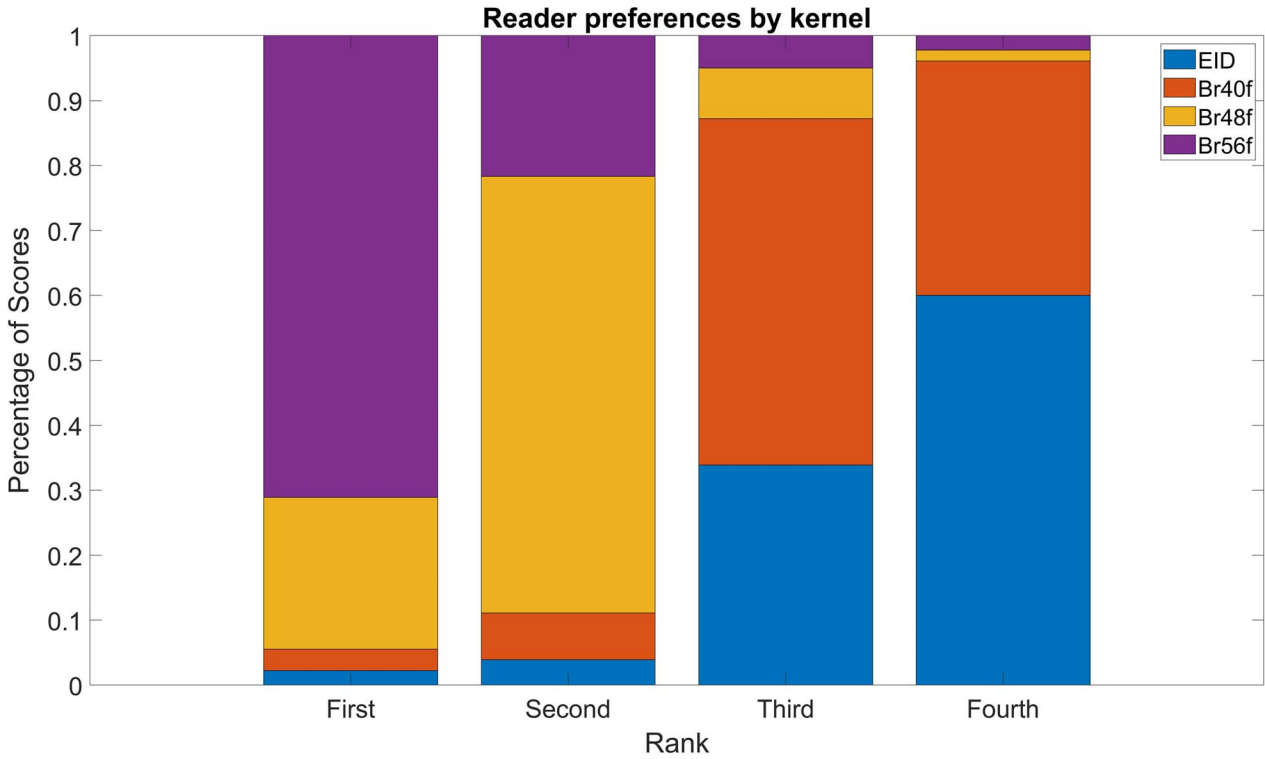


FIGURE 4. Overall scores for cases across the reader study. The x-axis represents different ranks, whereas the y-axis represents the percentage of scores receiving that rank. Each image source is indicated with a different color (EID, blue; Br40f, red; Br48f, yellow; Br56f, purple). EID indicates energy-integrating detector.

Br56f and Br48f had the fewest, with a combined 2% of all fourth ranks. Interrater reliability was high with a Shrout-Fleiss intraclass correlation coefficient of 0.990.

This was further reflected in a case-by-case breakdown by readers and patient cases (Fig. 5). Br56f was ranked first among 8 of 9 readers, whereas EID was ranked last among 6 of 9 readers. Similarly, Br56f was ranked first across 19 of 20 patients, and EID was ranked last across 16 of 20 patients. The Wilcoxon rank-sum test showed that the results for each pair of readers and patients were different ($P > 0.05$).

Average GNI for the EID CT images was 28.4 HU (range, 9.9–47.3 HU). For PCCT studies, average GNI was 21.2 HU (range, 13.4–29.1 HU), 34.6 HU (range, 21.6–48.1 HU), and 58.1 HU (range, 36.7–79.7 HU) for Br40f, Br48f, and Br56f kernels, respectively. Compared with EID CT, d' values for PCCT were found to be 15.9% higher for the Br40f kernel, 4.4% lower for the Br48f kernel, and 33.2% lower for the Br56f kernel images.

VIT Results

Across all simulated cases, changes in lung shape and volume during the scan had a noticeable impact on histograms of lung intensity, whereas changes in horizontal or angular displacement had a minimal effect (Fig. 6).

This was further reflected by the histogram intensity metrics that showed more variation due to changes in respiratory phase compared with horizontal or angular displacement (Table 4). Across the 3 kernels, Br40f had the lowest mean and standard deviation of absolute difference (mean [SD], 117 [133] HU) when compared with Br48f (mean [SD], 121 [134] HU) and Br56f (mean [SD], 139 [140] HU). The difference between Br40f and

Br48f was found to be statistically insignificant ($P = 0.22$), whereas the difference between Br56f and either Br40f or Br48f was found to be significant ($P < 0.01$).

Br56f showed lower bias than Br48f or Br40f for 13 of 17 features across both respiratory frame (Fig. 7) and displacement conditions (Fig. 8). In terms of variability, Br56f had slightly lower coefficients of variation across all radiomics features for both respiratory frame (Fig. 9) and displacement (Fig. 10) conditions.

DISCUSSION

Photon-counting CT is a technological development that can push the performance of CT beyond current clinical practice. In the context of thoracic imaging, this improvement is primarily due to improved spatial resolution and performance under noisy conditions. As the technology becomes more widely available, it becomes important to determine how to best use that technology for different clinical needs and to best integrate the systems within the existing clinical workflow. There are multiple methods to evaluate an imaging system for this purpose, which range from the physically informative, such as phantom-based methods, to the clinically grounded, namely, reader studies. In this work, we applied multiple approaches, namely, a task-specific phantom study, a clinically driven reader and in vivo characterization study, and a VIT, to assess the choice of reconstruction kernel for thoracic imaging in a clinical PCCT scanner.

The physical phantom approach allows for a direct characterization of the inherent image quality properties of the system. Using imaging phantoms with static features, the characterization of systems is robust and can be done over multiple acquisition parameters. However, these phantoms are limited in their

Downloaded from http://journals.lww.com/jcat by BhDM5fPHeKAvIzEoum11QINna+KLLHEZpsH04XNM0h0CwCk1AV nYQp/IOH-D3I3D000dRy/ITV5FACI3V/C1y0abggQZKdwmIKZBYws= on 04/17/2024

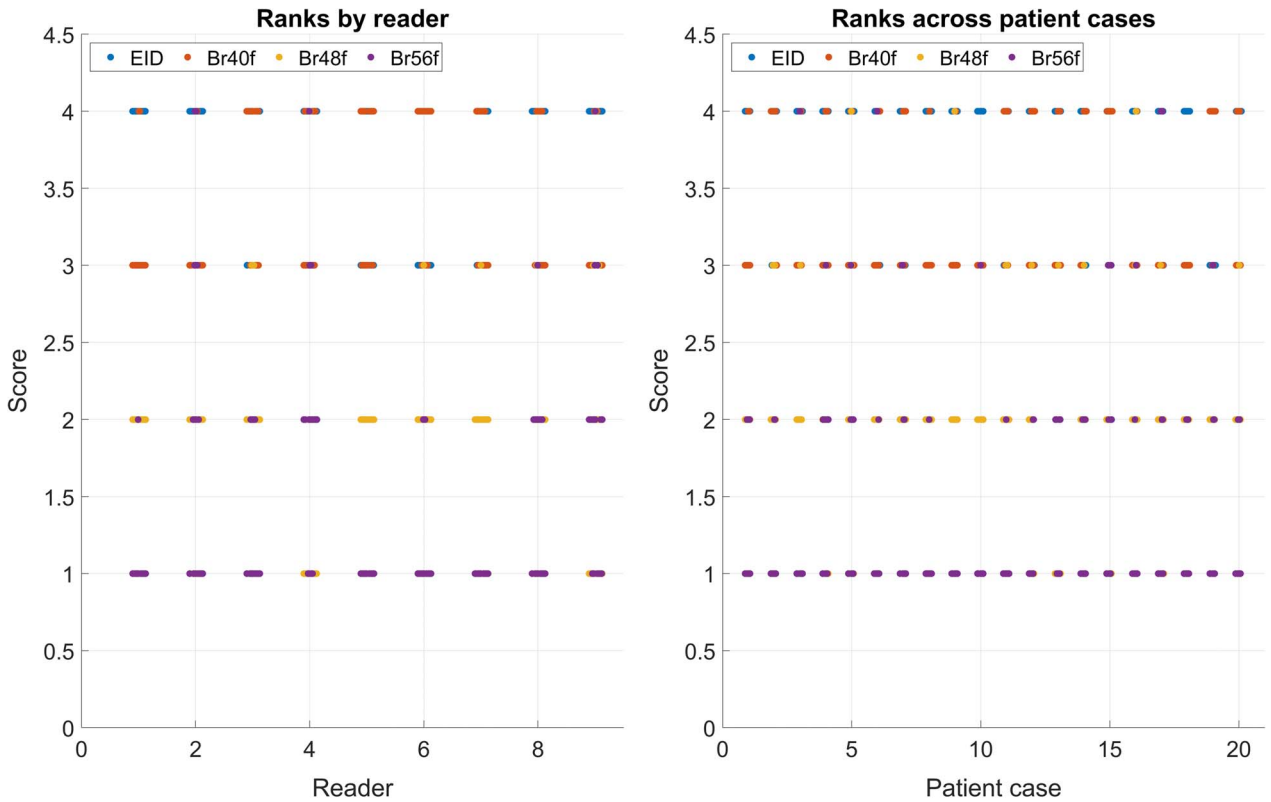


FIGURE 5. Chart displaying the ranks for each image kernel separated by reader preference (left) and patient case (right). Ranks range from 1 (least preferred) to 4 (most preferred). Each image source is indicated with a different color (EID, blue; Br40f, red; Br48f, yellow; Br56f, purple). EID indicates energy-integrating detector.

geometric complexity and clinical realism. When comparing across the 3 kernels evaluated in this study, Br40f was the least noisy but had the lowest spatial resolution. Conversely, Br56f was the noisiest with the highest spatial resolution. These findings translated into CNR and detectability index with Br40f

yielding a performance higher than that of Br56f. These measurements compare favorably with earlier studies: Bhattarai et al⁴⁶ evaluated the same PCCT and an EID CT system (Siemens Force) using 2 kernels, the softer of which was the same as our study (Br40) and the sharper of which (Br64) was

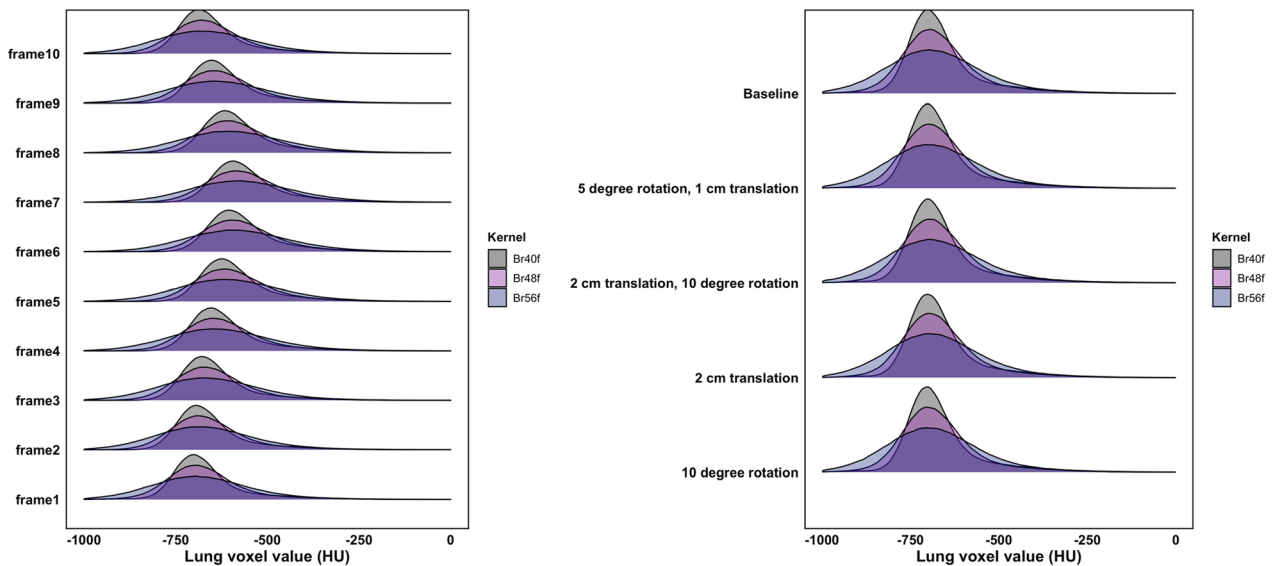


FIGURE 6. Density plots of lung intensity histograms for different respiratory phases of the simulated phantom (left) and different angular and positional displacements (right). Lung intensity (HU) is indicated along the x-axis, and specific case is indicated along the y-axis. Different kernels are indicated by color (Br40f, gray; Br48f, purple; Br56f, blue). HU indicates Hounsfield unit.

TABLE 4. Difference in Lung Density Histogram Measurements Due to Change in Condition

Condition	Br40f		Br48f		Br56f	
	Absolute Difference	Standard Deviation	Absolute Difference	Standard Deviation	Absolute Difference	Standard Deviation
Baseline	108.2	119.4	112.4	120.0	130.5	126.8
	Difference From Baseline	Difference From Baseline	Difference From Baseline	Difference From Baseline	Difference From Baseline	Difference From Baseline
Motion - frame 2	2.0	3.5	2.0	3.3	2.1	2.9
Motion - frame 3	5.8	10.1	5.9	9.9	6.1	9.0
Motion - frame 4	11.8	19.6	12.0	19.3	12.1	17.9
Motion - frame 5	18.1	28.6	18.4	28.4	18.3	26.5
Motion - frame 6	22.4	35.2	22.8	34.9	22.4	32.7
Motion - frame 7	25.2	39.1	25.5	38.8	25.2	36.4
Motion - frame 8	19.7	31.2	20.0	31.0	19.8	28.9
Motion - frame 9	11.7	19.6	11.9	19.4	12.0	18.0
Motion - frame 10	3.7	6.5	3.7	6.3	3.8	5.7
10° rotation	0.2	0.4	0.1	0.4	-0.3	0.2
5° rotation, 1-cm transition	0.2	0.3	0.2	0.3	0.1	0.2
2-cm transition	0.3	-0.1	0.4	0.0	0.7	0.2
10° rotation, 2-cm transition	1.0	0.2	1.5	0.4	2.3	0.8

Values are calculated as the absolute difference between ground truth and measured condition at the voxel level. Other conditions are reported as difference from the baseline.

sharper than the kernels we used (Br56). Bhattarai et al found that the EID system had a comparable range of noise f_{av} (0.25–0.28 mm^{-1} for Br40, 0.37–0.63 mm^{-1} for Br64) and TTF

f_{50} (0.30–0.39 mm^{-1} for Br40f, 0.55–1.09 mm^{-1} for Br64) values, whereas noise magnitude was found to have a larger range (2.2–51.54 HU for Br40, 18.9–127.5 HU for Br64). Although

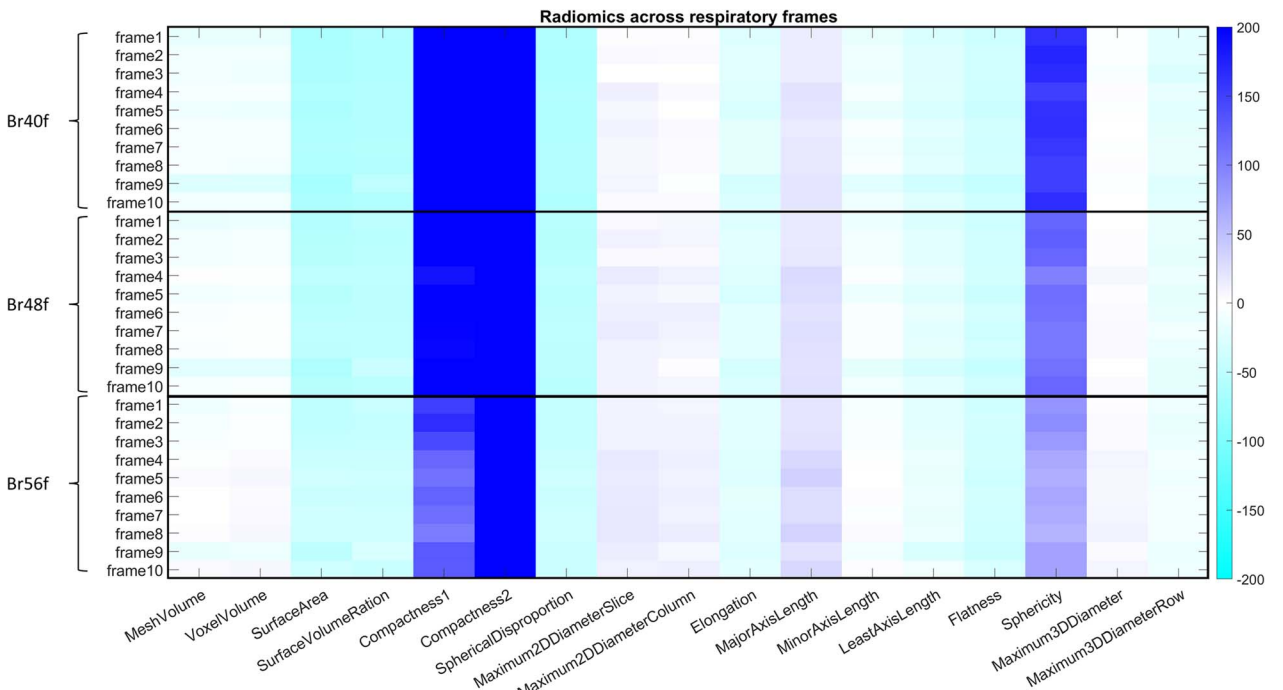


FIGURE 7. Percentage error for radiomics measures of lesions across respiratory phases. Radiomics measures are listed along the x-axis. Different conditions are separated along the y-axis, grouped by brackets to represent different kernel groupings. Percentage error is represented by dark blue for positive error and light blue for negative error.

Downloaded from http://journals.lww.com/jcat by BhDM5fPHeKAvIzEoum11QIN4a+KLiHEZqpsHh4XIM0h0ncYwCX1AW nYQp/IOH-H3I3D000dRyITV5FA#C3V/C1y0abggQZxdmwnkZB7w5= on 04/17/2024

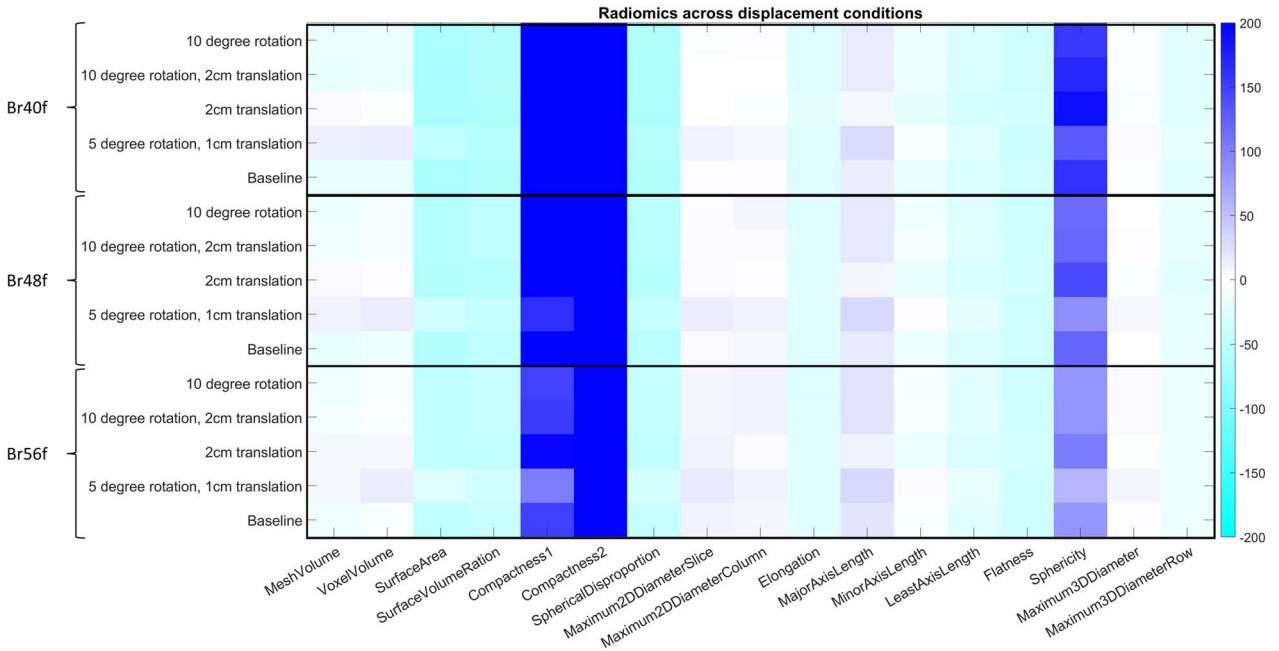


FIGURE 8. Percentage error for radiomics measures across displacement cases. Radiomics measures are listed along the x-axis. Different conditions are separated along the y-axis, grouped by brackets to represent different kernel groupings. Percentage error is represented by dark blue for positive error and light blue for negative error.

there were some differences in imaging protocol, including a different range of dose levels, the overall measurements were comparable between the 2 studies.

The clinical in vivo characterization corroborated the phantom evaluation as Br40f again had the lowest noise and highest detectability

performance whereas Br56f had the highest noise and lowest detectability. However, the clinical reader evaluation did not concord with these results based on noise alone, where radiologists indicated a clear preference for the Br56f kernel. When comparing across the 3 kernels evaluated in this study, Br56f was the most

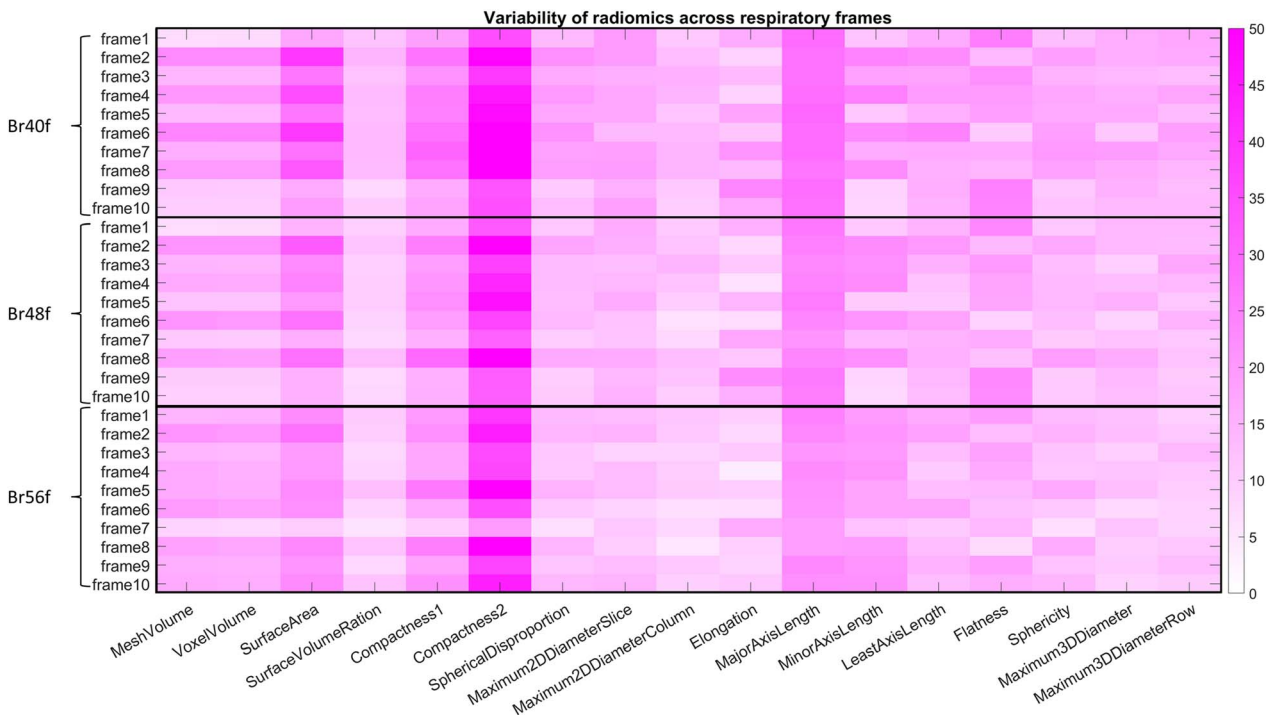


FIGURE 9. Coefficient of variation for radiomics measures of lesions across respiratory phases. Radiomics measures are listed along the x-axis. Different conditions are separated along the y-axis, grouped by brackets to represent different kernel groupings. Percentage error is represented by dark pink for positive error and light pink for negative error.

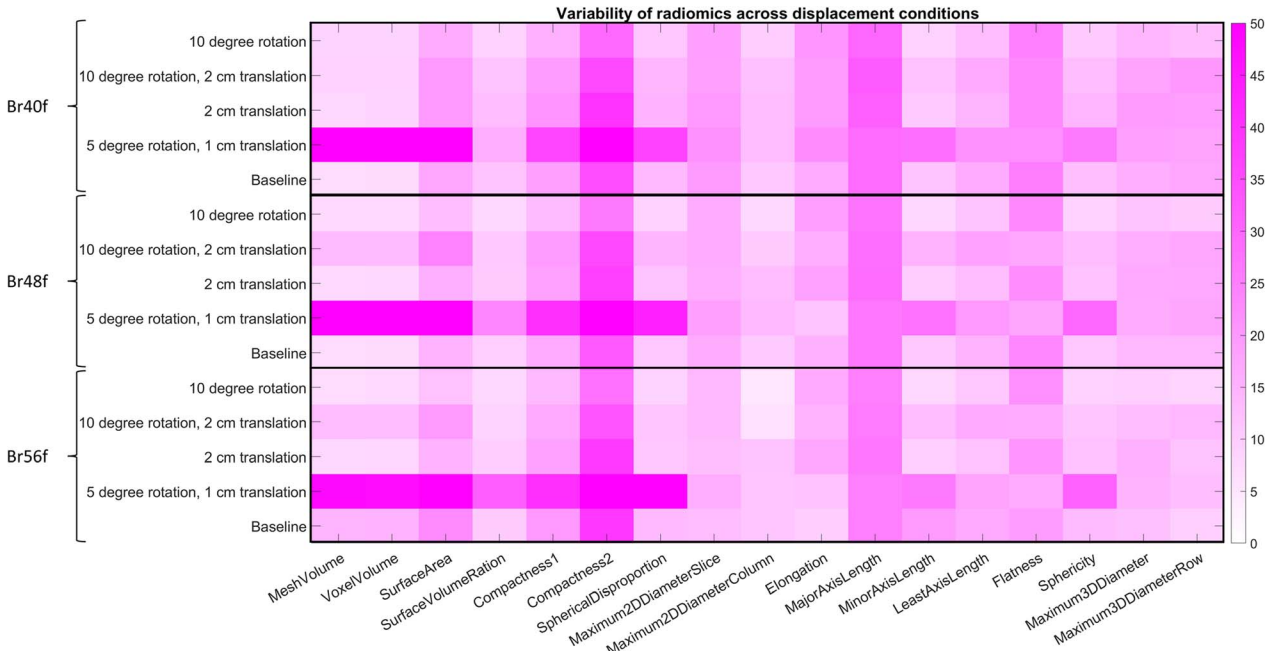


FIGURE 10. Coefficient of variation for radiomics measures across displacement cases. Radiomics measures are listed along the x-axis. Different conditions are separated along the y-axis, grouped by brackets to represent different kernel groupings. Percentage error is represented by dark pink for positive error and light pink for negative error.

preferred kernel across readers and across patient cases when viewed at lung windows. High intraclass correlation showed this preference to be consistent across institutions. Readers showed an overall preference for PCCT images over conventional EID images despite the PCCT images being acquired with a lower radiation dose, which shows that PCCT allows both image quality improvement and radiation dose reduction. This is presumably based on the nature of thoracic imaging tasks, focused on the visualization of subtle features of high-contrast structures. The best kernel for this task would be a kernel with a high spatial resolution and a reasonable noise level without as much a focus on CNR, best exemplified by Br56f among the kernels evaluated in this study.

Virtual imaging trials proved to offer a complementary perspective. Such trials allow for medical imaging experiments that are not possible with physical phantom or clinical studies. Because the patient being scanned is virtual, scans can be repeated with varying conditions for extensive task-based evaluations. Another benefit of VITs is the precise knowledge of spatial distribution and material properties of the object being imaged. Ground truth provides a precise point of comparison for evaluations that is unaffected by variability in physical measurements. In this study, we took advantage of both aspects of virtual clinical trials by exploring the effects of reconstruction kernels on lung density and lung lesion morphology compared with digitally defined, known ground-truth measures. We found that changes in respiratory phase caused a greater change in lung density than horizontal or angular displacement, as expected. For the lung density quantification analysis, Br40f had lower mean and standard deviation of absolute error (9.4 ± 14.9 HU) when compared with Br48f (9.6 ± 14.8 HU) and Br56f (9.6 ± 13.8 HU). For lung lesion morphology, Br56f had lower bias and variability than either Br40f or Br48f. The VIT results further indicated inpatient variability due to respiratory phase to be in the same order of magnitude as the influence due to kernel choice (3.3–38.1 HU in lung density and 9.9–21.4 voxels in lesion volume due to respiratory phase vs 19.8–21.5 HU in lung

density and -13.9 to 16.0 voxels in lesion volume due to the kernel choice). Previous VIT evaluations have also compared this PCCT scanner with clinical EID.⁴⁷ Sotoudeh-Paima et al²¹ compared EID and PCCT for the task of chronic obstructive pulmonary disease quantification and found PCCT to be superior. Likewise, Ho et al⁴⁸ found that PCCT provided superior quantification of bronchial airways. In both studies, a sharper kernel was found to provide the best quantification regardless of imaging system, mostly consistent with our VIT findings as well.

There have also been prior investigations that have used multiple methods to evaluate PCCT scanners focusing on different clinical tasks in cardiac,⁴⁹ abdominal,^{50,51} and neurological imaging.⁵² Within the context of thoracic imaging, a recent work⁵³ evaluated the same PCCT scanner used in this study to answer a similar question focusing on reduced dose imaging with iterative reconstruction. That study included a patient population data set evaluated using in vivo and reader evaluations. They found close concordance between the 2 methods. Although the specific questions differ between ours and theirs, they both demonstrate the utility of using multiple methods. Although no method is perfect, each method can provide a unique insight, that either is complementary with the results from another, thus affirming a particular solution, or highlights a perspective that needs to be balanced with other findings. This provides a more robust confidence for a final recommendation for protocols.

Overall, our study found the 4 evaluative methods each to provide complementary insight. The physical phantom evaluation found that the Br40f kernel had the least noise and lowest resolution whereas Br56f had the most noise and highest resolution. As a result, the Br40f kernel had the higher CNR and detectability index measurements. The in vivo clinical study further supported these findings. However, the reader study found that the Br56f kernel was most preferred by readers, suggesting that for this task, the benefit in resolution outweighed the increased noise and lower CNR or detectability measurements. The VIT showed that Br56f

Downloaded from http://journals.lww.com/jcat by BhDM5fepPkKav1zEoum1tQIN4a+KLLhEzpsH64XW0h0CwCX1AV nYQp/IOH3D3D00dRyITVfACI3V/C1y0abggQZkXdmwKZB7wS= on 04/17/2024

likewise provided a more quantitative depiction of focal lesions in terms of both bias and variability. However, for the task of lung density quantification, a softer kernel yielded more precise and accurate estimates, and those findings were consistent with the physical phantom and in vivo image quality studies. Unique to VIT methodology, the results further illustrated the magnitude of inpatient variability and the importance of its mitigation (eg, consistent level of inspiration during a test) for quantitative imaging tasks.

In this study, we deployed 4 distinct methods to assess CT systems. As illustrated, each evaluation method provides insight on a different aspect of a system performance. Not every site would be able to use each method and evaluate all necessary protocol questions due to cost or logistical limitations. Yet, knowing what each method can offer, pros and cons as, and limitations, a user can make a better-informed choice based on the specific clinical question at hand. In our study, radiologists' preference did not match the detectability metric results from the phantom study. We attribute the difference to the fact that detectability index does account for some but not all factors of relevance to image perception. Theoretically, the metric provides a representation of the information content of the image with respect to signal and distracting noise based on an assumed task. The real task might differ from what was assumed, and furthermore, there is certainly more to human cognition than information content, because of the perpetual efforts to improve observer models. Furthermore, the reader evaluation in this study was a ranking by preference between different image types and not signal detection. If the readers were instead asked to perform a signal detection task, we would have expected a closer correlation with the detectability index findings, as demonstrated in prior studies.

There were some limitations to this study. The specific methods used in this study, in terms of physical phantom choice, reader study design, and virtual models, represent a small subset of all possible study designs. Other tasks would require future studies. Additionally, this study assessed the question of which kernel was best for a specific clinical task. Further optimization of other parameters would require additional evaluations. Our virtual trial was also limited to only 1 virtual patient. A larger cohort of patients would offer systematic assessment of variability across patients.

CONCLUSION

Although there are different methods for determining which protocol design is the best for any given clinical task, there are complementary values to each. Physical phantom studies characterize physical attributes of imaging systems but can be limited in their clinical applicability. In vivo and reader studies using clinical data provide information about clinical needs and preferences but can be limited by sample size of both patient and reader populations and thus may suffer from lack of statistical power. Virtual imaging trials enable task-specific evaluations with repeated measures and knowledge of ground truth, enabling assessment of both quantitative performance, and interpatient and inpatient variability. In this work, we found that for the task of determining which kernel to use for a general thoracic scanning protocol with a PCCT system, all 4 methods can provide essential and effectual perspectives, and the best kernel for use was dependent on clinical task. Aggregating our findings, we concluded the Br56f kernel best for high-resolution tasks and Br40f for contrast-dependent tasks.

ACKNOWLEDGMENTS

The authors would like to thank Travis Henry, Kevin Kalisz, Lynne Koweek, Ashkan Malayeri, Bryan O'Sullivan-Murphy, Babak Saboury, and Arlene Sirajuddin for their participation as

readers in the observer study. They acknowledge support from Siemens Healthineers for this project.

REFERENCES

1. Ria F, Solomon JB, Wilson JM, et al. Technical note: validation of TG 233 phantom methodology to characterize noise and dose in patient CT data. *Med Phys*. 2020;47:1633–1639.
2. Abadi E, Segars WP, Tsui BM, et al. Virtual clinical trials in medical imaging: a review. *J Med Imaging (Bellingham)*. 2020;7:042805.
3. Samei E, Abadi E, Kapadia A, Lo J, Mazurowski M, Segars P. Virtual imaging trials: an emerging experimental paradigm in imaging research and practice. Paper presented at: SPIE Medical Imaging 2020, February 20, 2020, Houston, TX. *Physics of Medical Imaging*; 2020.
4. Webb WR, Muller NL, Naidich DP. *High-Resolution CT of the Lung*. Lippincott Williams & Wilkins; 2014.
5. Yanagawa M, Johkoh T, Noguchi M, et al. Radiological prediction of tumor invasiveness of lung adenocarcinoma on thin-section CT. *Medicine (Baltimore)*. 2017;96:e6331.
6. Islam N, Ebrahimzadeh S, Salameh J-P, et al. Thoracic imaging tests for the diagnosis of COVID-19. *Cochrane Database Syst Rev*. 2021;3:CD013639.
7. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax*. 2012;67:296–301.
8. Çorbacioğlu SK, Er E, Aslan S, et al. The significance of routine thoracic computed tomography in patients with blunt chest trauma. *Injury*. 2015;46:849–853.
9. Rubinowitz AN, Siegel MD, Tocino I. Thoracic imaging in the ICU. *Crit Care Clin*. 2007;23:539–573.
10. Just KS, Defosse JM, Grensemann J, et al. Computed tomography for the identification of a potential infectious source in critically ill surgical patients. *J Crit Care*. 2015;30:386–389.
11. Awerbuch E, Benavides M, Gershengorn HB. The impact of computed tomography of the chest on the management of patients in a medical intensive care unit. *J Intensive Care Med*. 2015;30:505–511.
12. Taguchi K, Iwaczyk JS. Vision 20/20: single photon counting x-ray detectors in medical imaging. *Med Phys*. 2013;40:100901.
13. Farhadi F, Rajagopal JR, Nikpanah M, et al. Review of technical advancements and clinical applications of photon-counting computed tomography in imaging of the thorax. *J Thorac Imaging*. 2021;36:84–94.
14. Rajagopal JR, Sahbaee P, Farhadi F, et al. A clinically driven task-based comparison of photon counting and conventional energy integrating CT for soft tissue, vascular, and high-resolution tasks. *IEEE Trans Radiat Plasma Med Sci*. 2021;5:588–595.
15. Rajagopal JR, Farhadi F, Solomon J, et al. Comparison of low dose performance of photon-counting and energy integrating CT. *Acad Radiol*. 2021;28:1754–1760.
16. Gutjahr R, Halaweish AF, Yu Z, et al. Human imaging with photon counting-based computed tomography at clinical dose levels: contrast-to-noise ratio and cadaver studies. *Invest Radiol*. 2016;51:421–429.
17. Leng S, Zhou W, Yu Z, et al. Spectral performance of a whole-body research photon counting detector CT: quantitative accuracy in derived image sets. *Phys Med Biol*. 2017;62:7216–7232.
18. Symons R, Pourmorteza A, Sandfort V, et al. Feasibility of dose-reduced chest CT with photon-counting detectors: initial results in humans. *Radiology*. 2017;285:980–989.
19. Symons R, Cork TE, Sahbaee P, et al. Low-dose lung cancer screening with photon-counting CT: a feasibility study. *Phys Med Biol*. 2017;62:202–213.
20. Zhou W, Montoya J, Gutjahr R, et al. Lung nodule volume quantification and shape differentiation with an ultra-high resolution technique on a

- photon-counting detector computed tomography system. *J Med Imaging (Bellingham)*. 2017;4:043502.
21. Sotoudeh-Paima S, Segars WP, Samei E, Abadi E. Photon-counting CT versus conventional CT for COPD quantifications: intra-scanner optimization and inter-scanner assessments using virtual imaging trials. Paper presented at: SPIE Medical Imaging 2022, February 24, 2022, San Diego, CA. Physics of Medical Imaging; 2022.
 22. US Food and Drug Administration. FDA clears first major imaging device advancement for computed tomography in nearly a decade [press release], September 30, 2021. Silver Spring, MD: US Food and Drug Administration; 2021. Available at: <https://www.fda.gov/news-events/press-announcements/fda-clears-first-major-imaging-device-advancement-computed-tomography-nearly-decade>. Accessed August 2, 2023.
 23. Rajendran K, Petersilka M, Henning A, et al. First clinical photon-counting detector CT system: technical evaluation. *Radiology*. 2022;303:130–138.
 24. Solomon J, Zhang Y, Wilson J, Samei E. An automated software tool for task-based image quality assessment and matching in clinical CT using the TG-233 framework. Paper presented at: the AAPM Annual Meeting 2018, August 2, 2018, Nashville, TN.
 25. Samei E, Bakalyar D, Boedeker KL, et al. Performance evaluation of computed tomography systems: summary of AAPM Task Group 233. *Med Phys*. 2019;46:e735–e756.
 26. Boedeker KL, Cooper VN, McNitt-Gray MF. Application of the noise power spectrum in modern diagnostic MDCT: part I. Measurement of noise power spectra and noise equivalent quanta. *Phys Med Biol*. 2007;52:4027–4046.
 27. Siewerdsen JH, Cunningham IA, Jaffray DA. A framework for noise-power spectrum analysis of multidimensional images. *Med Phys*. 2002;29:2655–2671.
 28. Chen B, Christianson O, Wilson JM, et al. Assessment of volumetric noise and resolution performance for linear and nonlinear CT reconstruction methods. *Med Phys*. 2014;41:071909.
 29. Robins M, Solomon J, Richards T, et al. 3D task-transfer function representation of the signal transfer properties of low-contrast lesions in FBP- and iterative-reconstructed CT. *Med Phys*. 2018;45:4977–4985.
 30. Richard S, Husarik DB, Yadava G, et al. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. *Med Phys*. 2012;39:4115–4122.
 31. Solomon J, Wilson J, Samei E. Characteristic image quality of a third generation dual-source MDCT scanner: noise, resolution, and detectability. *Med Phys*. 2015;42:4941–4953.
 32. ICRU. *Report 54: Medical Imaging—The Assessment of Image Quality*. Bethesda, MD: International Commission on Radiation Units and Measurements; 1995.
 33. Chen B, Richard S, Christianson O, Zhou X, Samei E. CT performance as a variable function of resolution, noise, and task property for iterative reconstructions. Paper presented at: SPIE Medical Imaging 2012, February 9, 2012, San Diego, CA. Physics of Medical Imaging; 2012.
 34. Salarian A. Intraclass correlation coefficient (ICC). In *MATLAB Central File Exchange*. 2022. Available at: <https://www.mathworks.com/matlabcentral/fileexchange/22099-intraclass-correlation-coefficient-icc>. Accessed February 13, 2022.
 35. Sanders J, Hurwitz L, Samei E. Patient-specific quantification of image quality: an automated method for measuring spatial resolution in clinical CT images. *Med Phys*. 2016;43:5330–5338.
 36. Smith TB, Solomon JB, Samei E. Estimating detectability index in vivo: development and validation of an automated methodology. *J Med Imaging (Bellingham)*. 2018;5:031403.
 37. Cheng Y, Abadi E, Smith TB, et al. Validation of algorithmic CT image quality metrics with preferences of radiologists. *Med Phys*. 2019;46:4837–4846.
 38. Abadi E, Harrawood B, Sharma S, et al. DukeSim: a realistic, rapid, and scanner-specific simulation framework in computed tomography. *IEEE Trans Med Imaging*. 2019;38:1457–1465.
 39. Jadick G, Abadi E, Harrawood B, et al. A scanner-specific framework for simulating CT images with tube current modulation. *Phys Med Biol*. 2021;66:185010.
 40. Abadi E, Harrawood B, Rajagopal JR, et al. Development of a scanner-specific simulation framework for photon-counting computed tomography. *Biomed Phys Eng Express*. 2019;5:055008.
 41. Segars WP, Bond J, Frush J, et al. Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. *Med Phys*. 2013;40:043701.
 42. Sauer TJ, Samei E. Modeling dynamic, nutrient-access-based lesion progression using stochastic processes. Paper presented at: SPIE Medical Imaging 2019, February 21, 2019, San Diego, CA. Physics of Medical Imaging; 2019.
 43. McCabe C, Zarei M, Segars WP, Samei E, Abadi E. Optimization of imaging parameters of an investigational photon-counting CT prototype for lung lesion radiomics. Paper presented at: SPIE Medical Imaging 2022, February 24, 2022, San Diego, CA. Computer-Aided Diagnosis; 2022.
 44. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Process*. 2001;10:266–277.
 45. Van Griethuysen JMM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107.
 46. Bhattarai M, Bache S, Abadi E, et al. A systematic task-based image quality assessment of photon-counting and energy integrating CT as a function of reconstruction kernel and phantom size. *Med Phys*. 2024;51:1047–1060.
 47. McCabe C, Sauer TJ, Zarei M, Segars WP, Samei E, Abadi E. A systematic assessment of photon-counting CT for bone mineral density and microarchitecture quantifications. Paper presented at: SPIE Medical Imaging 2023, February 23, 2023, San Diego, CA. Physics of Medical Imaging; 2023.
 48. Ho FC, Sotoudeh-Paima S, Segars WP, Samei E, Abadi E. Development and application of a virtual imaging trial framework for airway quantifications via CT. Paper presented at: SPIE Medical Imaging 2023, February 23, 2023, San Diego, CA. Physics of Medical Imaging; 2023.
 49. Rajagopal JR, Farhadi F, Richards T, et al. Evaluation of coronary plaques and stents with conventional and photon-counting CT: benefits of high-resolution photon-counting CT. *Radiol Cardiothorac Imaging*. 2021;3:e210102.
 50. Sartoretto T, Landsmann A, Nakhostin D, et al. Quantum iterative reconstruction for abdominal photon-counting detector CT improves image quality. *Radiology*. 2022;303:339–348.
 51. Sartoretto T, Mergen V, Higashigaito K, et al. Virtual noncontrast imaging of the liver using photon-counting detector computed tomography: a systematic phantom and patient study. *Invest Radiol*. 2022;57:488–493.
 52. Pourmorteza A, Symons R, Reich DS, et al. Photon-counting CT of the brain: in vivo human results and image-quality assessment. *AJNR Am J Neuroradiol*. 2017;38:2257–2263.
 53. Sartoretto T, Racine D, Mergen V, et al. Quantum iterative reconstruction for low-dose ultra-high-resolution photon-counting detector CT of the lung. *Diagnostics (Basel)*. 2022;12:522.