AAPM REPORT NO. 233

Performance Evaluation of Computed Tomography Systems

The Report of AAPM Task Group 233

April 2019

DISCLAIMER: This publication is based on sources and information believed to be reliable, but the AAPM, the authors, and the editors disclaim any warranty or liability based on or relating to the contents of this publication.

The AAPM does not endorse any products, manufacturers, or suppliers. Nothing in this publication should be interpreted as implying such endorsement.

© 2019 by American Association of Physicists in Medicine

This page intentionally left blank.

Performance Evaluation of Computed Tomography Systems

The Report of AAPM Task Group 233

Ehsan Samei[1], Donovan Bakalyar[2], Kirsten L Boedeker[3], Samuel Brady[4], Jiahua Fan[5], Shuai Leng[6], Kyle J. Myers[7], Lucretiu M. Popescu[8], Juan Carlos Ramirez Giraldo[9], Frank Ranallo[10], Justin Solomon[11], Jay Vaishnav[12], and Jia Wang[13]

Grateful Acknowledgment:

Nicholas Bevins
Guang-Hong Chen
Dianna Cody
Eric Gingold
Loretta Johnson
David Jordan
Xiang Li
Jeffrey Limmer
Cynthia McCollough
Michael McNitt-Gray
Christina Skourou

Disclosures:

Jiahua Fan is an employee of GE Healthcare. Lucretiu Popescu is an employee of Neusoft Medical Systems. Kirsten Boedeker and Jay Vaishnav are employees of Canon. Juan Carlos Ramirez Giraldo is an employee of Siemens Healthineers. Ehsan Samei, Duke University, has a nonexclusive license agreement with the Sun Nuclear Corporation in commercializing the Mercury Phantom. His proceeds from the sale of the phantom are donated to the AAPM Research and Education fund.

[1,11]Duke University
[2]Henry Ford Health System
[3,12]Canon Medical Systems
[4]Cincinnati Children's Hospital
[5]GE Healthcare
[6]Mayo Clinic
[7]FDA
[8]Neusoft Medical Systems
[9]Siemens Healthineers
[10]University of Wisconsin
[13]Stanford University

DISCLAIMER: This publication is based on sources and information believed to be reliable, but the AAPM, the authors, and the publisher disclaim any warranty or liability based on or relating to the contents of this publication.

The AAPM does not endorse any products, manufacturers, or suppliers. Nothing in this publication should be interpreted as implying such endorsement.

ISBN: 978-1-936366-69-9
ISSN: 0271-7344

© 2019 by American Association of Physicists in Medicine

All rights reserved

Published by

American Association of Physicists in Medicine
1631 Prince Street
Alexandria, VA 22314

Contents

Introduction

The rapid development and complexity of new x-ray computed tomography (CT) technologies, the increased utilization of CT, and the need for evidence-based optimization of image quality with respect to radiation and contrast media dose call for an updated approach to evaluating the performance of CT systems. In light of the availability and increasing clinical use of new CT technologies, it is particularly important to assess image quality using task-specific metrics that are more relevant to predicting the performance of a CT system and protocols for clinical imaging tasks.

A prevalent new CT technology uses statistical and iterative reconstruction (IR) algorithms to decrease image noise to facilitate use of decreased radiation dose levels. The nonlinear nature of these algorithms results in object-dependent resolution and noise performances. Thus, traditional image quality metrics, such as contrast-to-noise ratio, have become inadequate indicators of clinical imaging performance. While such traditional image quality indicators retain their relevance for evaluation of CT equipment technical performance, they fall short as surrogates of clinical performance for either product evaluation or optimization purposes. Furthermore, automatic exposure (AEC) techniques, such as tube current modulation (TCM) techniques, have become ubiquitous in the clinical practice of CT. Methods are needed to characterize the performance of TCM techniques to better inform users as to how the system's radiation output is adapted to patient attributes.

This document aims to supplement and complement existing and prior equipment performance testing guidelines (e.g., AAPM Report 74[1]) by addressing the more advanced aspects of current CT systems, such as IR and TCM. The goal of this report is to briefly summarize current performance evaluation metrics and quality control (QC) tests, and introduce advanced performance assessment methods within a single document.* Pass-fail criteria or performance guidelines are not provided for the results of these advanced assessment methods; there are no manufacturer specifications or regulatory or accreditation performance requirements available for these quantities. Rather, in line with the current professional trajectory of the field toward operational engagement, it is hoped that the assessment methods described in this report will be adopted by the clinical medical physicist for the purposes of protocol optimization, and for indicating clinical imaging performance in a way that can be compared between systems and imaging protocols. These important assessment methods also pave the way to approach performance testing of new CT systems, not only in terms of acceptance testing (i.e., verifying a device meets predefined specifications), but also for system commissioning (i.e., determining how the system can be used most effectively in clinical practice).

The primary audience for this report is medical physicists working in the field of CT, especially those physicists involved in clinical performance optimization, but we envision that this report can provide informative definitions of assessment methods for anyone with an interest in CT performance evaluation. Throughout this report, the terms "physicist" and "clinical physicist" are used synonymously with "qualified medical physicist" (QMP), defined by the AAPM as an individual who is competent to independently provide clinical professional services in one or more subfields of medical physics (AAPM Professional/Education/Science Policy 1). In this report, it is assumed that a QMP has competency in diagnostic medical physics.

---

* Each manufacturer provides its own recommendations for QC procedures in accordance with international and regulatory standards. This report is meant to supplement and not to replace such procedures. Adherence to manufacturer-recommended QC procedures for specific CT systems facilitates communication between the vendor, service engineer, and user. Further, the information provided herein is not a formal dissemination of information by FDA and does not represent Agency position, policy, or guidance for industry.

This report is organized into three main sections as detailed in Table 1. Sections 1 and 2 provide a summary of established techniques for characterizing the basic performance of CT systems. The primary focus of these sections is to offer a concise and unified outline of methodologies that are already well described in the associated references. Its second aim is to provide a comprehensive resource on CT characterization and QC testing to allow readers to gain a broad perspective on CT performance testing without the need to move across multiple references.

Tests described in sections 1 and 2 are oriented toward the technical performance of the CT system and are appropriate for ensuring that the equipment meets manufacturer specifications, as well as regulatory and accreditation requirements.

In contrast, section 3 targets operational performance of a CT system and clinical protocols, with metrics that more directly reflect clinical performance. These aspects of CT performance are not yet fully established and, thus, the section is primarily descriptive; no pass/fail criteria or guidelines are given, as the clinical performance requirements depend not only on the specific clinical task and patient characteristics, but also on the expectations and preferences of the specific clinical practice.

Rather, the assessment methods outlined in section 3 are oriented toward characterizing the CT system not in terms of its technical specifications (e.g., tube potential accuracy, exposure linearity), but in terms that are largely independent of the specific technologies implemented by a manufacturer. As such, they are oriented toward assessing system performance in a manner that is a better indicator of its clinical utility.

Table 1. Components of CT performance evaluation as structured in this report

Performance Type	Performance Sub-type (Section Number)	Component
Pre-test inspection	Basic functional and QC (1)	Specific checks prior to basic and operational tests
Basic performance	Geometrical performance (2.1)	Laser alignment accuracy
		Table indexing accuracy
		Image position accuracy
		Image thickness accuracy (axial mode)
		Image thickness accuracy (helical mode)
		Gantry tilt accuracy
	Radiation output performance (2.2)	Half-value layer
		Exposure reproducibility
		Exposure time reproducibility
		Exposure linearity
		Exposure time accuracy
		Tube potential accuracy
		Radiation beam profile
		Displayed $CTDI_{vol}$ accuracy
		CT localizer radiograph dose
	Basic imaging performance (2.3)	CT number accuracy
		CT number uniformity
		Artifact assessment
		Line-pair resolution
		Noise magnitude
		Slice sensitivity profile
Operational performance	Advanced imaging performance (3.1–3.3)	Tube current modulation
		Spatial resolution
		Noise
	Task-based performance (3.4–3.5)	Quasi-linear task-based performance
		Spatial domain task-based performance

Finally, the characterization of a medical imaging system is most meaningful to the extent that it is predictive of the clinical outcome (Figure 1). The premise of this report is that while metrology summarized in sections 1 and 2 evaluates technical conformance to established specifications, the metrology detailed in section 3 can be more readily related to the clinical outcome, and thus can serve as a more meaningful surrogate of CT performance in patient care tasks, hence the term operational performance.

Many of the quantities and their associated assessment methodologies given in section 3 of this report reflect work done by the imaging research community. The hope and premise of this report is that these quantities can be used by clinical physicists to assess operational performance. To reach this goal, consistent specific definitions for these quantities must be established and tools must be made publicly available to make such measurements practical in a clinical environment. This report aims to address both of these obstacles so as to foster their clinical use as a way to make the physics characterizations of CT more scientifically informed and clinically relevant.

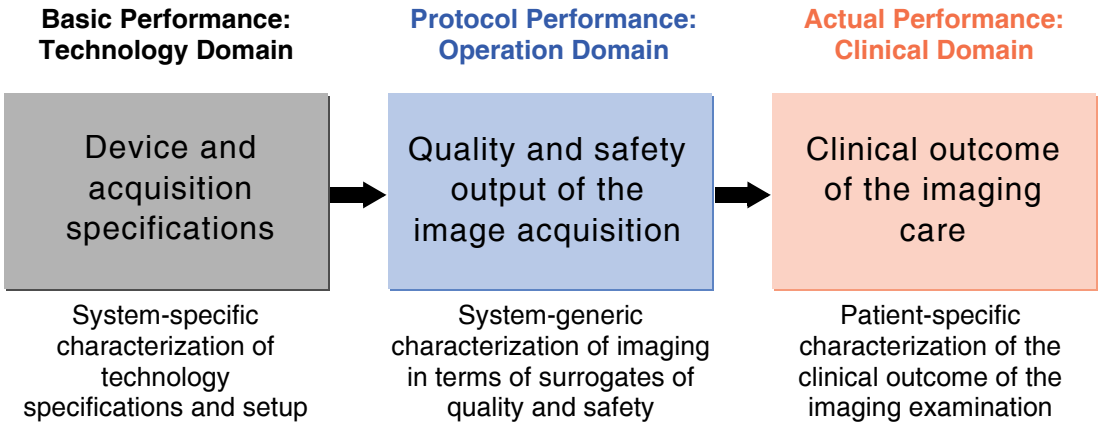

Figure 1. Components of CT performance evaluation in the technology domain (mainly section 2 of this report) and operational domain (mainly section 3 of this report).

1. Pre-test Inspection and QC Program Review

A pre-test inspection for the overall state of a CT system is an important prerequisite of CT system evaluation, particularly for a newly installed system. This inspection is not technically a component of the physics performance evaluation, but is required nonetheless in advance of any follow-up basic or advanced testing procedures. Some pre-test inspections are listed in Table 2. Such tests should be informed by local, state, and national regulations and accreditation policies.

Table 2. Elements of pre-test inspection of CT systems

	Task	Detail
1	"Caution Radiation Area" warning sign at all entrances to the scan room and other required postings	Verify presence of warning signs if required by your state regulations or institutional policy. Also check for the presence of other postings required by your state, accreditation bodies, and institution. Note that the code of federal regulations (CFR) regarding radiation warning signs (10 CFR 20.1901–1905) from the Nuclear Regulatory Commission (NRC) are intended to apply only to radioactive materials licensed by the NRC and thus do not directly apply to radiation-producing machines. However, individual states may have posting requirements for radiation-producing machines such as CT scanners. Independent of regulatory requirements, such postings are recommended by this task group at all entrances to the CT scan room.
2	Functioning of x-ray indication light at room entry	If you have x-ray indication lights, confirm their proper operation. Consult your state regulations or institutional policy to determine if they are required, and comply with regulations and policies.
3	Doors to room closed when making an exposure	If the room has doors, confirm their proper operation. Consult your state regulations or institutional policy to determine if they are required, and comply with regulations and policies. Note: This can create difficulties when transporting patients through the doors, requiring "workarounds" from the staff.
4	Labeling, visibility and access to emergency stops	Emergency stops must be present, operable, and accessible. This last point is extremely important. For example, wall-mounted emergency stops can inadvertently be made inaccessible by a moving cart or by rearranging the room. See CFR rule 21CFR1020.33.
5	X-ray warning label at the control panel	See CFR rule 21CFR1020.30 (j): "Warning label. The control panel containing the main power switch shall bear the warning statement, legible and accessible to view: "Warning: This x-ray unit may be dangerous to patient and operator unless safe exposure factors, operating instructions and maintenance schedules are observed."
6	Production of a clearly noticeable signal during x-ray generation	This signal can be visible, audible, or both. See CFR rule 21CFR1020.
7	Direct line of sight to patient during procedure	Operator should have a direct line of sight to patient during CT examination. See CFR rule 21CFR1020.
8	Visual determination of a reference tomographic plane or reference plane offset	See CFR rule 33, 21CFR1020.33 (g). For CT, the rule reads as follows: "For any multiple tomogram system, means shall be provided to permit visual determination of the location of a reference plane. The relationship of the reference plane to the planes of the tomograms shall be provided to the user in addition to other information provided according to 1020.30(h). This reference plane can be offset from the location of the tomographic planes." The physicist should confirm adherence to this rule. If lasers are used, they must be usable under ambient lighting conditions of 500 lux.
9	Operator initiation of scans	Initialization of exposure must require positive, deliberate action by the operator. See Canada's Safety Code 35: "Safety Procedures for the Installation, Use and Control of X-ray Equipment in Large Medical Radiological Facilities."
10	Oral communication between operator and patient	Operator should be able to orally communicate with the patient from within the control room. See CFR rule 21CFR1020.
11	Display of technique factors before scan	The anticipated technique factors (e.g., kV, mAs) must be clearly indicated to the operator prior to the actual scan. See CFR rule 21CFR1020.
12	Pre-scan display of prescribed $CTDI_{vol}$, DLP, and size of phantom	The pre-scan (i.e., anticipated) $CTDI_{vol}$ and DLP should be displayed to the operator. The size of the CTDI phantom (i.e., 32 cm diameter or 16 cm diameter) should be indicated.[2,3]
13	Post-scan $CTDI_{vol}$, DLP and phantom size recording	The post-scan (i.e., delivered) $CTDI_{vol}$ and DLP should be displayed in the patient's dose report.[2,3]

(continued)

14	Operator and service manuals present	Verify presence of operator, service, and other necessary technical manuals provided by the manufacturer. See CFR rule 21CFR1020.
15	MITA Smart Dose	National Electrical Manufacturers Association (NEMA) XR-29 Standard[4], also known as MITA Smart Dose, requires CT scanners to incorporate automatic exposure control (NEMA XR-28-2013), DICOM-compliant radiation dose structured reporting (NEMA XR-29-2013), dose check features (NEMA XR-25-2010[5]), and reference pediatric and adult protocols (NEMA XR-28-2013). Each manufacturer has a vendor certification web portal on the MITA Smart Dose website (http://www.medicalimaging.org/policy-and-positions/mita-smart-dose/). Medical physicists should contact the manufacturer of each CT scanner to upgrade software, if needed, and to obtain a verification of compliance. Without compliance (effective Jan. 1, 2016), CMS reimbursement is reduced by 5% for outpatient scans in 2016 and by 15% in 2017. Note that compliance with XR-29 is established based on CT scanner capabilities, not on usage. When used appropriately, these features improve the safety of CT scans for the patients and can improve the image quality.
16	Dose data connectivity	Following MITA Smart Dose provision above, the scanner should ideally not only be capable of producing a DICOM radiation dose structured report, but also send the report to a destination for interpretation (e.g., via a dose monitoring system). This should ideally be integrated with a system by which the institution's technologists and radiologists properly understand the use and interpretation of dose results, and dose notification and alert values in accordance with institutional policies.
17	Presence of adequate shielding	If a new room has been constructed for your machine, a medical physicist, radiation safety officer, or other qualified personnel should visually confirm the presence of shielding during construction, whenever possible. The best direct assessment technique uses a radioactive source and detector (in accordance with appropriate safety measures). An alternative technique is to use a portable x-ray source and detector. A third technique to confirm the presence of adequate shielding is to use the scattered radiation emitted when scanning a phantom (e.g., the body CTDI phantom) and a high-sensitivity radiation measurement device. If a new machine is being installed, a medical physicist, radiation safety officer, or other qualified personnel should check the integrity of the shielding, since this could have been compromised during the construction process. One of the techniques outlined above should be used. It is imperative to ensure compliance with state regulations, which may involve design and verification of shielding integrity by a physicist and communicating the results of that verification to a state agency.[6]
18	Presence of an appropriate QC program	A continuous quality control program (QC) is required[7] as part of CT accreditation programs (e.g., ACR, IAC, TJC) and may also be required by other accreditation bodies, the state, or by the institution. The presence and appropriate implementation for such a program should be verified as part of this pre-test inspection, and if a QC program does not exist or is insufficient, one should be properly instituted.

2. Basic System Performance

A number of foundational characteristics govern the inherent performance of CT systems. They can be broadly characterized into three basic aspects of the system: geometric accuracy, radiation output, and imaging output. It is imperative that such characteristics are evaluated by a clinical physicist at commissioning (prior to first clinical use) and throughout the life of a CT system. The methods to ascertain such basic system characteristics are well established by several noteworthy organizations as listed below.

1. American Association of Physicists in Medicine (AAPM) Report 74[8] and Report 39[9]
2. American College of Radiology (ACR) CT Quality Control (QC) Manual for the ACR CT Accreditation Program (CTAP)[7]
3. European Commission (EC) Report 162 on CT quality assurance[10]

4. Food and Drug Administration Code of Federal Regulations (CFR) part 120[11]
5. International Atomic Energy Agency (IAEA) Human Health Series No. 19[12] and Human Health Report No. 5[13]
6. International Commission on Radiation Units & Measurements (ICRU) Report 87[14]
7. International Electrotechnical Commission (IEC) report 61223-3-5[15]
8. Vendor quality control documents

In this section, we briefly describe the purpose of each of these basic testing methods succinctly, with some further details presented in tabular form in the appendix. For complete descriptions of each test, readers are encouraged to seek the original source material. As far as this report is concerned, for a given CT system, exactly what attribute of the CT performance should be characterized with what testing method is up to the discretion of the physicist. In this section, the aim of this report is only to provide a summary of existing tests defined by *other* resources.

It is important to note that manufacturers' quality control programs are an important component of CT performance as they are designed with the specific design of each specific CT system's hardware and software in mind, in accordance with IEC. Service and other supporting personnel are trained according to the manufacturer's procedures and, thus, are best able to assist users when these manufacturer-provided QC procedures are followed. In addition, failure to maintain the system using the QC protocols, procedures, and frequencies recommended by the manufacturer may result in voiding of warranties and may be in violation of a site's purchase or service agreement. Alternative testing should be used in addition to, not in place of, manufacturer-provided QC.

2.1 Geometric Performance

Geometrical performance pertains to the basic aspects of the system functioning that are related to spatial reproducibility and accuracy. The purpose of each test is listed below, and descriptions of the tests are given in the appendix in tabular form (section 6.1, Table 5).

- **Laser alignment accuracy**: To ensure that the laser alignment lights correctly indicate the scan position.
- **Table indexing accuracy**: To ensure that the table moves as indicated.
- **Image position accuracy**: To ensure that the prescribed image location indicated in a CT localizer radiograph correctly corresponds to the image position
- **Image thickness accuracy**: To ensure that the nominal reconstructed image thickness is similar to the actual reconstructed image thickness
- **Gantry tilt accuracy**: To ensure that the nominal gantry tilt is similar to the actual gantry tilt and that the gantry returns to a vertical position after being tilted

2.2 Radiation Output Performance

Radiation performance pertains to characterization of the radiation output of the CT system. The purpose of each test is listed below and descriptions of the tests are given in the appendix in tabular form (section 6.2, Table 5).

- **Half-value layer**: To measure the half-value layer of the CT system's x-ray source and ensure that it is within regulatory limits
- **Exposure reproducibility**: To ensure the radiation output of the system is consistent across repeated identical exposures

- **Exposure time reproducibility**: To ensure the exposure time is consistent across repeated identical exposures
- **Exposure linearity**: To ensure the radiation output of the system is linearly proportional to mAs
- **Exposure time accuracy**: To ensure the nominal exposure time is similar to the actual exposure time
- **Tube potential accuracy**: To ensure the nominal tube potential is similar to the actual tube potential
- **Radiation beam profile**: To ensure the nominal radiation beam width is similar to the actual beam width
- **Displayed $CTDI_{vol}$ accuracy**: To ensure the displayed $CTDI_{vol}$ is similar to the actual $CTDI_{vol}$
- **CT localizer radiograph dose**: To measure the exposure from the localizer radiograph

2.3 Basic Image Quality Performance

Image quality performance pertains to the aspects of system performance that are related to characterization of reconstructed images. The purpose of each test is listed below, and descriptions of the tests are given in the appendix in tabular form (section 6.3, Table 6).

- **CT number accuracy**: To ensure the CT numbers reported by the scanner are within an acceptable tolerance for known materials
- **CT number uniformity**: To ensure acceptable uniformity in CT numbers across the image field of view
- **Artifact assessment**: To ensure the images are free from artifacts.
- **Line-pair (high-contrast) resolution**: To estimate the limiting high-contrast (in-plane) spatial resolution of the system
- **Noise magnitude**: To characterize the first-order noise properties of the CT system and to ensure the noise is consistent over time
- **Low-contrast contrast-to-noise ratio (CNR)**: To estimate the low-contrast performance of the CT system and ensure that it is acceptable for diagnosis
- **Slice sensitivity profile (SSP)**: To estimate the high-contrast z-direction spatial resolution of the system

3. Operational Performance

The basic system characteristics summarized in the prior section reflect the intrinsic performance of a CT system. While those characteristics provide a first-order depiction of a system's functionality, they do not reflect a number of features and attributes of CT systems that affect the quality of patient images. In this report, such attributes are recognized under the heading of Operational Performance. Operational performance characterization of a CT system aims to provide a metrology more closely reflective of performance of the system in clinical imaging. Operational performance characterization should thus provide a stronger clinical basis by which to evaluate the system's performance, and supports further use of the measurements for optimizing the system for targeted image quality or dosimetric goals. Furthermore, this section describes several system characterization methodologies designed to assess important CT adaptive technologies that have been introduced to reduce radiation dose and optimize image quality (e.g., TCM or IR) for which established physics testing methods (in sections 1 and 2) are not well suited to address.

The manner in which these characterization assessment methods could or should be applied in a clinical context is still emerging; adaptation of these methods, for example, for regulatory compliance or accreditation purposes, would require considerable additional effort before clinical adoption could be required. However, the ubiquity of newer CT technologies, such as TCM and IR, and their large impact on image quality, mandate that the clinical medical physicist be involved in their characterization, implementation, and optimization. Thus, this section aims to provide a common "toolset" that a clinical physicist can use for characterization and optimization purposes. Pass/fail criteria are not provided for these testing methods as such data are not yet available based on peer performance or concordance with clinical outcome data. Further, such data are expected to be application- and radiologist-specific. The approach of this section is a departure from traditional conformance/specification based (i.e., pass-fail) physics testing. The idea is not to pass or fail a system based on these measurements, but rather to use these measurements to improve the understanding and utilization of the technology.

For example, task-based performance measurements (sections 3.4 and 3.5) could be used as the basis for establishing a protocol in terms of reconstruction kernel, image thickness, and radiation output when implementing IR into a clinical protocol. Additionally, task-based performance measurements can be monitored and tracked across a clinical operation, paving the way for image quality registries and standards of practice (see also section 4). Table 3 lists the quantitative metrics described in this report, with the descriptions and mathematical definitions of each metric described in subsequent sections.

The philosophy and metrology of operational performance aims to make the evaluation more reflective of clinical performance with the use of phantoms that offer greater variability, as would be expected in real patients (e.g., phantoms that reflect variations in patient size), and more technology-relevant methods (e.g., testing methods that accommodate potential system nonlinearities). However, that goal cannot be perfectly achieved with phantom test objects. For example, patient habitus, limb position, weight, and tissue heterogeneities can all affect patient image quality. In many cases, the adaptive nature of modern CT systems can be invoked under certain conditions to accommodate such patient heterogeneities. However, while some technologies share basic principles across CT scanners (e.g., TCM or IR), their implementation, user interface, user options, and conditions of use may vary substantially across manufacturers and even across models within a manufacturer. Therefore, a thorough review of the manufacturers' documentation, testing conditions, and conditions of use is extremely important.

While operational characterization is shown to be more closely correlated with clinical performance[16–19] than basic characterization, this correlation has not been demonstrated for all imaging tasks and metrics. Additional efforts are needed to determine the performance of current clinical CT systems based on the metrics and methodologies described in this report and, further, to fully ascertain the quantitative dependencies of clinical performance on these metrics. These efforts will pave the way to establishing performance targets and tolerances for each metric.

Table 3. Metrics of CT operational performance

Attribute	Section	Metric	Definition
Tube Current Modulation	3.1	g_{mA}	Functional dependence of tube current on water equivalent diameter for a given phantom
		g_n	Functional dependence of noise on water equivalent diameter for a given phantom
		α, R_A	Exponent and the correlation coefficient of ln(mA) = $\alpha(d_w)+\beta$ relationship for a given phantom
		s, R_n	Slope and the correlation coefficient of n = $s(d_w)+t$ relationship for a given phantom
		C_{mA}, C_{noise}	Spatial concordance, in mm, of the distance between a discontinuous change in thickness and the anticipated change in mA or noise
		d_{min}, d_{max}	Diameters associated with mA of the system reaching its maximum or its minimum value
Spatial Resolution	3.2	$TTF_{n,C}$	Task Transfer Function (TTF) at defined measured noise and contrast level in the in-plane direction
		$zTTF_{n,C}$	Task Transfer Function (TTF) at defined measured noise and contrast level in the z-direction (i.e., trans-axial direction)
		f_{50} and f_{10}	Frequencies associated with 50% and 10% of in-plane TTF, respectively
		zf_{50} and zf_{10}	Frequencies associated with 50% and 10% of z-direction TTF, respectively
Noise	3.3	n	Noise magnitude (pixel standard deviation) at three dose levels
		NPS_n	Noise power spectrum (NPS) at defined noise levels
		f_P and f_A	Peak and average frequencies of the NPS
		NUI	Noise nonuniformity index
		η	Noise inhomogeneity index
Quasi Task-Based Performance	3.4	d'	Detectability index for the detection of a target signal (e.g., 1, 5, and 10 mm circular signal having a specific contrast and contrast-profile) for a specific phantom size and noise or dose level
		e'	Estimability index for estimating the volume of a target signal (e.g., 10 mm spherical signal having a specific contrast and contrast-profile) for a specific phantom size and noise or dose level
Spatial Domain Task-Based Performance	3.5	LR	Localization success rate for identifying the presence of and location of a targeted signal
		A_{LROC}	Area under the localization relative operating characteristic (LROC) curve for targeted localization tasks
		A_{EFROC}	Area under the exponential transformed free response operating characteristic (EFROC) curve for targeted free-response detection tasks

The tests involved in this report require imaging specific types of phantoms; some appropriate phantoms are suggested in this report. The imaging can be done using any protocols that the user may wish to evaluate. Typically, these tests should be performed under sample conditions of interest representative of the protocol and dose conditions used or to be used clinically (e.g., typical head and body protocols as common reflections of clinical operation). However, comparing tests conducted across different systems is possible if a common protocol is used. A set of suggested testing protocols are thus listed in Table 4. These protocols aim to provide an overall broad characterization of the system, in lieu of or in addition to any specific protocol(s) of interest that the user may wish to evaluate.

Table 4. A suggested list of testing protocols for task-based characterization of CT systems. Use an explicit typical FOV appropriate for the phantom and typical beam collimation width (e.g., 40 mm).

Nomenclature*	CTDI (mGy) (32 cm Phantom)	Tube Potential (kV)	Tube Current (mA)	Mode, Pitch	Reconstruction
TG233-F1	0.75	120	Fixed mA to achieve target CTDI ±10%	Helical, ~1	FBP, IR at medium strength, higher than medium strength, and maximum strength settings "standard" kernel ~0.6 and 5 mm image thickness
TG233-F2	1.5				
TG233-F3	3.0				
TG233-F4	6.0				
TG233-F5	12.0				
TG233-F6	24.0				
TG233-F3LK	3.0	70 (or 80)		Helical, ~1, unless a lower pitch is needed to achieve the CTDI	
TG233-F3MK	3.0	100			
TG233-F3HH	3.0	150 (or 140)			
TG233-M2	1.5	120	TCM setting to achieve target CTDI ±10%	Helical, ~1	
TG233-M3	3.0				
TG233-M4	6.0				
TG233-M3-A	3.0	120	Same as above	Axial	

* F refers to fixed mA, M to TCM, 1–6 to dose setting, and LK, MK, HK to low, medium, and high kV settings, respectively.

3.1 Tube Current Modulation

3.1.1 Objective

To characterize the tube current modulation (TCM) in terms of tube current and image noise as a function of attenuation. Two complementary tests are presented: One assesses how a CT system adapts the tube current to a discrete change in object attenuation and size, and the other how it does so in response to a continuous change.

3.1.2 Important Definitions

- *Tube current*: It determines the number of electrons accelerated across the x-ray tube per unit time. It is expressed in units of milliAmperes (mA). The CT scanner radiation output, in terms of $CTDI_{vol}$, is directly proportional to the tube current.
- *Tube current modulation (TCM):* This scanner feature automatically adapts the x-ray tube current to the patient attenuation to achieve a specified level of image quality. Most modern CT systems can modulate the tube current in several directions (see angular and longitudinal modulation below) or synchronized with an ECG signal.
- *Automatic exposure control (AEC):* Any system that automatically adapts the tube output (e.g., tube current, tube potential, etc.) according to the radiological properties of the patient. Technically, TCM is a specific implementation or type of AEC. However, in the literature, AEC and TCM are sometimes used synonymously.
- *Angular modulation of the tube current (x-y modulation):* This TCM feature adapts the tube current as the x-ray tube rotates around the patient to compensate for attenuation changes at varying projection angles, attempting to control detector signal at different projection angles. The angular modulation usually uses one or two CT localizer radiographs (in some systems in combination with the detector signal from prior rotations) to estimate patient attenuation.

- *Longitudinal modulation of the tube current (z-modulation):* This TCM feature adapts the tube current as patient attenuation changes in the longitudinal direction. The longitudinal modulation usually uses one or two CT localizer radiographs to estimate patient attenuation.

3.1.3 Equipment

Various sets of phantoms can be used for this procedure, depending on how complete of a characterization is sought. The *discrete adaptation test* utilizes a phantom of different fixed sizes (at least two) in the longitudinal direction. The *continuous adaptation test* uses a phantom with continuous changes in water-equivalent diameter in the longitudinal direction. A phantom may also be used that includes a combination of both discrete and continuous changes in size. Sections 5.2.2 and 5.2.3 show examples of available phantoms.

3.1.4 Procedures

The objective of this investigation is to assess how the CT system adapts the tube current as a function of object size with either discrete or continuous changes in attenuation under a fixed set of operating conditions.

For either test, start by defining a set of operating conditions according to scanner model and manufacturer for a routine adult body protocol using 120 kV. Scan in helical mode with pitch of ~1.0, and rotation time of 1 s (Table 4, TG233-M2, M3, or M4). Alternatively use a sequential (axial) mode with rotation time of 1 s (Table 4, TG233-M3-A). Use default TCM settings according to scanner model and manufacturer (i.e., noise index, standard deviation, quality reference mAs, etc.). Additional protocols may be used to ascertain the sizes at which the mA of the system maxes out to its highest value or bottoms down to its lowest, both of which can change as a function of the kV and the phantom size.

For the discrete adaptation test, scan at least two different-sized phantoms, each centered precisely, to assess the amount of tube current adaptation. Prior to each scan, perform one or two CT localizer radiographs covering the full range of the phantom, according to manufacturer recommendation (i.e., "AP" or "AP + lateral" directions). It is important to note that the order in which the CT localizer radiographs are obtained can affect the resulting TCM profile for certain systems. In such systems, often the final CT localizer radiograph is used for TCM prescription. Define a CT scan range that starts and ends at least half of the total collimation away from both edges of the phantom, otherwise the air boundary of the phantom will impact the results. Reconstruct the images using a standard body kernel. Reconstructing with thin slices (<1 mm) is preferred as it will provide a finely sampled mA profile, but at the cost of many more images to reconstruct, transfer, and process. Thus, thicker slices could be used but at the cost of potentially losing details in the extracted mA profile (see next section).

For the continuous adaptation test, use a phantom with continuously varied size or attenuation, position the phantom at isocenter, and perform one or two CT localizer radiographs over the full range of the phantom, according to manufacturer recommendation (i.e., "AP" or "AP + lateral" directions). Define a CT scan range that starts and ends at least half of the total collimation away from both edges of the phantom. Reconstruct the images using a standard body kernel with an image thickness of 5 mm with an interval of 5 mm.

3.1.5 Data Analysis

For the discrete adaptation test, use either the scan protocol page or the resulting images to record the mA (or mAs) and $CTDI_{vol}$ values of each of the CT scans performed with the different-sized phantoms (Figure 2). Note that mA or mAs per image is a scalar (often average) representation of the tube current, which can vary as a function of tube position. Also the $CTDI_{vol}$ can vary throughout the scan, so the scanner reported value is also an averaged value. Trace a circular region of interest (ROI) at the

	Scan	kV	mAs / ref.	CTDIvol* mGy	DLP mGycm	Tl s	cSL mm
Topogram	1	120	35 mA	0.13 L	10	7.8	0.6
CTDI32	2	120	215 / 210	14.48 L	124	1.0	0.6
Topogram	1	120	35 mA	0.13 L	4	3.2	0.6
CTDI16	2	120	43 / 210	2.90 L	25	1.0	0.6
Topogram	1	120	35 mA	0.13 L	5	3.9	0.6
CTDI10	2	120	23 / 210	1.55 L	13	1.0	0.6

Figure 2. Example of size adaptation test of the TCM. Images were acquired with a Siemens Definition AS64 scanner using adult body protocol, 120 kV, rotation time = 1s, and pitch = 1.0. For this specific scanner and manufacturer, the TCM (CAREDose4D) was set with 210 quality reference mAs, with curves setting at 'average.' Three CTDI phantoms of size 32, 16, and 10 cm were scanned independently using the same CT technique described above. Prior to each CT scan, a CT localizer radiograph was acquired in the anteroposterior (AP) direction. The scan protocol page above shows the tube current values were adapted to 215, 43, and 23 mA for the 32, 16, and 10 cm CTDI phantoms, respectively. In Siemens CT systems, rather than reporting 'mA,' the system reports effective mAs, which is defined as the tube current time product divided by the pitch. Because rotation time and pitch were (conveniently) set to 1, in this special case, the effective mAs equals the mA values.

center of the phantom, and measure the standard deviation of the CT numbers. ROI should have a diameter of 1 cm or more. Repeat for three contiguous images near the center of the scan, and report the average of the standard deviation for each of the phantom sizes.

For the continuous adaptation test, the tube current values can be extracted from the DICOM header (tag 0018,1151) of the CT images for each image position. These tube current values in a DICOM header typically represent the average tube current over all tube positions that contributed to that image. Trace a circular region of interest (ROI) (of at least 1 cm in diameter) at the center of the images, and measure the standard deviation of the CT numbers. Report the overall average of the measured image noise as a function of phantom size. Figure 3 shows an example.

For either test, with the size known at each position, apply a log-linear fit to mA versus phantom size, d_w (water equivalent diameter), as $\ln(\text{mA}) = \alpha(d_w)+\beta$, and report the slope α and the linear correlation coefficient R_{mA}. Apply a linear fit to the average of the measured standard deviation (n) versus phantom size (d_w) as $n = s(d_w)+t$, and report the slope s and the correlation coefficient R_n. See Figure 4 for an example. This analysis may also include ascertaining the sizes at which the mA of the system maxes out to its highest value or bottoms down to its lowest.

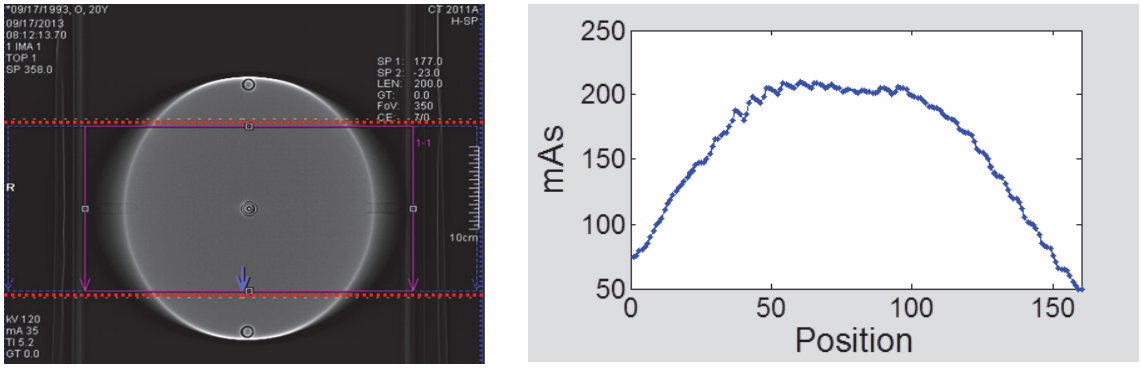

Figure 3. Example of the continuous adaptation test of the TCM using a 32 cm $CTDI_{vol}$ phantom placed with one of the flat cross-sectional ends of the phantom on the table surface. (Left) Displays the anteroposterior (AP) CT localizer radiographs, with dotted lines in AP radiographs, indicating the CT scan range. (Right) The tube current time product values (in units of mAs) are plotted as a function of z-axis position for the $CTDI_{vol}$ phantom. Data were collected with a single set of reference operating conditions using a Siemens Somatom Definition AS64 scanner using adult body protocol, 120 kV, rotation time of 1 s, and pitch of 1.0. The scanner-specific TCM (CAREDose4D) was set with 210 quality reference mAs with curves set to 'average.' These data demonstrate that the TCM system adjusts the tube current continuously as the attenuation changes continuously.

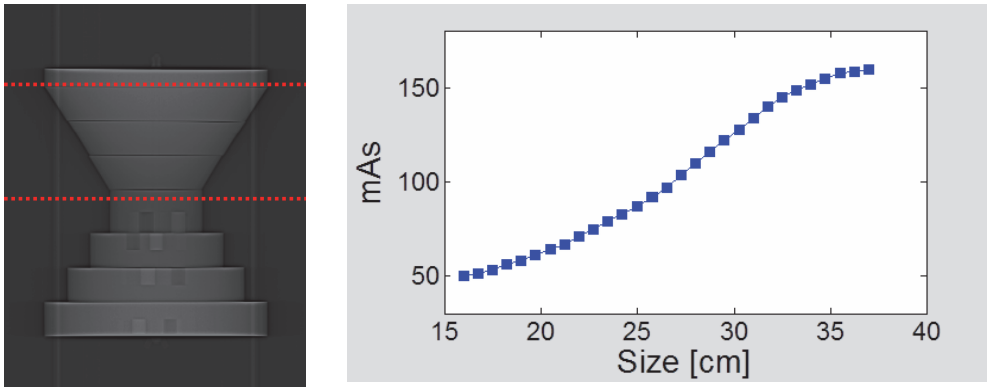

Figure 4. Example of the continuous adaptation test of the TCM using a tapered phantom (Mercury phantom, reconfigured for this test). (Left) Displays the anteroposterior (AP) CT localizer radiographs, with dotted lines in AP radiographs, indicating the CT scan range. (Right) The tube current time product values (in units of mAs) are plotted as a function of size. Data were collected with a single set of reference operating conditions using a Siemens Somatom Definition AS64 scanner using adult body protocol, 120 kV, rotation time of 1 s, and pitch of 1.0. The scanner-specific TCM (CAREDose4D) was set with 210 quality reference mAs with curves set to 'average.' These data demonstrate that the TCM system adjusts the tube current continuously as the attenuation changes continuously.

TG-233 recommends that size be described in terms of water-equivalent diameter. Water-equivalent diameter can be estimated using the methods described in AAPM Report 220.[20] The analysis for the discrete adaptation test should further include Spatial Concordance (C_{mA} or C_{noise}, the distance between a discontinuous change in phantom size, and the corresponding change in mA or noise, Figure 5).

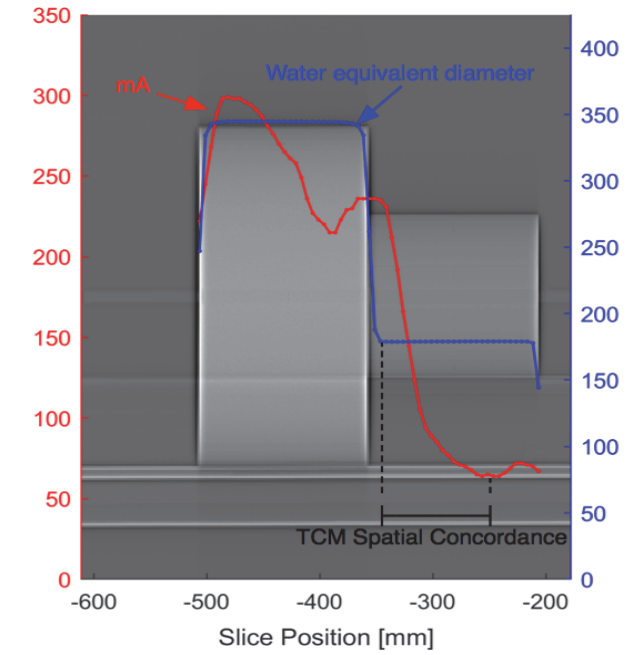

Figure 5. Example of the spatial concordance, C, between a discontinuous change in phantom size and the corresponding change in mA (or noise). In this example, two CTDI phantoms were set up side-by-side and scanned using TCM. The scout image from those scans is shown with the mA (red) and water-equivalent diameter (blue) shown for each slice position from the subsequent scans. The spatial concordance quantifies how "quickly" the CT system can adapt the tube current in concordance with an abrupt change in attenuation. In other words, it quantifies the lag between a change in attenuation and the responsive change in tube current. A spatial concordance of zero would imply perfect adaption to changing patient size. It would be expected that wider x-ray beam collimation settings would correspond to a larger (i.e., poorer) spatial concordance. This example demonstrates the spatial concordance of mA, C_{mA}. The spatial concordance of noise, C_{noise}, could be measured in a similar fashion, substituting the mA profile for the slice-by-slice profile of measured noise. It may also be possible to estimate the spatial concordance using a continuously changing phantom; a comparison between different phantom types for doing this test has not yet been made.

3.1.6 Precautions and Caveats

Select a scan range 'inside' the phantom in order to avoid imaging at edges along the z axis. As a rule of thumb, scan half a beam width inside of each edge of the phantom, as otherwise the air boundary of the phantom will impact the results. Note, however, that in some cases it may actually be of value to set the scan range over the ends of the phantom to observe the expected TCM behavior for very abrupt air-to-tissue interfaces (e.g., end of head or feet). For very large or very small phantoms, it is possible that no modulation occurs depending on the TCM settings. Note that the system mA maxing out to its highest value or bottoming down to its lowest can either be dictated by the system limitation, which can change as a function of the kV and the phantom size, or by the user protocol definition. Some TCM implementations allow the user to select minimum and maximum settings and, thus, these settings may need to be adjusted in order to observe normal TCM behavior in such phantoms.

3.1.7 Recommended Performance Metrics

- Functional dependence of mA on water equivalent diameter, $mA(d_w)$, for a given phantom
- Functional dependence of noise on water equivalent diameter, $n(d_w)$, for a given phantom

- Slope α and the correlation coefficient (R_{mA}) of ln(mA) = α(d_w)+β relationship (g_{mA}), for a given phantom
- Slope s and the correlation coefficient (R_n) of n=s(d_w)+t relationship (g_n), for a given phantom
- Spatial concordance, *C*, (in mm) of mA and noise change with discontinuous changes in size, the distance between a discontinuous change in phantom size, and the corresponding change in mA or noise (C_{mA} and C_{noise})
- Diameters associated with mA of the system reaching its maximum or its minimum value (d_{min} and d_{max})

3.1.8 References

- MHRA Report 0501[21]
- Solomon et al. 2015[22]
- Wilson et al. 2013[23]
- ICRU Report 87[14]

3.2 Spatial Resolution

3.2.1 Objective

To characterize the in-plane and z-axis spatial resolution of the CT system under reference conditions and establish baseline values for specific imaging conditions with methods applicable to both linear and nonlinear reconstruction algorithms. Note that the methods in this section are suitable to assess the spatial resolution for low-contrast features. As noted below, the traditional methods to assess spatial resolution using high-contrast line-pair patterns (see section 2.3) may not faithfully reflect the system's ability to resolve low-contrast features if nonlinear processing (e.g., iterative reconstruction) is used.

3.2.2 Important Definitions

- *Point Spread Function (PSF):* the system output response to an input point object
- *Line Spread Function (LSF):* the system output response to an input line object
- *Edge Spread Function (ESF):* the system output response to an input edge (i.e., step) object
- *Modulation Transfer Function (MTF):* Fourier transform (magnitude) of the LSF (normalized by the DC component). The MTF is a metric of system resolution and describes the system contrast recovery as a function of spatial frequency. This formulation assumes a linear, shift-invariant (LSI) imaging system. Although a rigorous mathematical description of an LSI system is beyond the scope of this report, practically speaking, an LSI system is one whose output signal can be determined by convolving an input signal with the system's PSF, independent of the properties or location of the input (e.g., contrast, size, central vs. peripheral).
- *Task Transfer Function (TTF):* the quasi-linear analog to the MTF. When the imaging system is known to behave nonlinearly (e.g., in the case of iterative reconstruction), a measured MTF may not represent the imaging system's response to an arbitrary input object (as would be the case for a truly linear system). In this scenario, the system resolution becomes dependent on the object contrast and background noise level. Therefore, the MTF is not general but rather "task-specific" and is denoted as a TTF. A TTF is measured in identical fashion as an MTF. However, when reporting a TTF, the background noise (pixel standard deviation, SD), object's contrast, and object's radial location should be included. Denoting the MTF as the TTF emphasizes that the system resolution is influenced by those factors. This emphasis becomes important when

computing task-based performance (see section 3.4) or when comparing resolution between different imaging systems or conditions. Traditionally, an MTF is measured using high-contrast objects, while a TTF should be measured for objects of a contrast that represents the imaging task under study.

3.2.3 Equipment

- For in-plane spatial resolution, a phantom with circular insert rods of varying contrast such as the Mercury phantom or the CT ACR 464 phantom
- For z-axis spatial resolution, a phantom in which two sections of different materials interface to form an edge parallel to the axial plane (e.g., interface between modules 2 and 3 of the CT ACR 464 phantom)
- Image analysis software capable of MTF calculations, see section 5.3

3.2.4 Test Procedures

3.2.4.1 In-plane Resolution

Align the phantom on the scanner. Make sure the phantom rods are perpendicular to the image plane. Image the phantom under sample conditions of interest representative of the protocol and dose conditions used or planned clinically or protocols defined in Table 4. To achieve a reliable estimate of the TTF, the total effective CNR (CNR you would achieve from averaging all available images), CNR_T, should be at least 15 according to Chen et al.[24] CNR_T can be computed as

$$CNR_T = CNR_S \times \sqrt{N}, \quad \text{(Eq. 1)}$$

where CNR_S is the CNR measured in an individual image and N is the number of images in which the rod is visible. Alternatively, the number of images needed to achieve a CNR_T of 15 can be computed as

$$N = \left(\frac{15}{CNR_S} \right)^2. \quad \text{(Eq. 2)}$$

Acquire the necessary number of images to achieve this CNR_T threshold. Note that this may involve repeated scans for low-contrast and low-dose conditions.

3.2.4.2 Z-axis Resolution

Align the phantom such that the axial-plane interface is slightly angled (approximately 5°) with respect to the image plane. With the ACR phantom, this can be achieved by adjusting the screw in its support base. It's also possible to use the gantry tilt feature, if available. Image the phantom under all conditions of interest representative of the protocol and dose conditions used or planned clinically. Figure 6 illustrates this setup.

3.2.5 Data Analysis

3.2.5.1 In-plane Resolution

Export the images to image analysis software to perform the TTF calculations (e.g., imQuest, section 5.3). The calculations should follow the circular rod method introduced by Richard et al.[25] and further refined by Chen et al.[24,22] In this method, a circular ROI with a radius about twice that of the phantom rod is roughly centered about the rod. The exact center location of the rod is estimated in each image

by finding the maximum of the cross-correlation between the image data and an idealized image of the phantom rod at upsampled resolution (to allow higher precision in center identification).

The precise center location of the rod is estimated for each image, and each pixel's radial distance from the center is calculated. An ESF is then generated by binning and averaging pixel CT numbers as a function of radial distance. Radial bin widths of $1/10^{th}$ the image pixel size are recommended. The derivative of the ESF is taken to yield the LSF. The TTF is finally computed as the magnitude of discrete Fourier transform of the LSF (normalized by the DC component).

Once the TTF is computed, the 50% and 10% frequencies (f_{50} and f_{10}) can be determined and used to summarize the system resolution under the given acquisition/reconstruction conditions. It is also important to report the contrast and noise conditions under which the TTF was measured. This analysis is illustrated in Figure 6.

3.2.5.2 Z-axis and 3D Resolution

Export the images to image analysis software to perform the TTF calculations (e.g., imQuest, section 5.3). The calculations should follow the slanted edge plane method described by Chen et al.[24] In this method, a virtual 2D plane is fit to the volumetric image data to determine the precise location and angle of the phantom plane edge. Based on this fit, a raw ESF is generated by plotting voxel intensity against signed (i.e., positive and negative) distance from this plane. The z-direction TTF is then calculated from this raw ESF using the same methods described above for the in-plane TTF. As a first approximation, the z-direction TTF is here assumed to be independent from the in-plane TTF to form a 3D TTF, i.e., $TTF(u,v,w) = TTF(u,v) \times TTF(w)$.

Once the TTF is computed, the frequencies associated with 50% and 10% TTF (f_{50} and f_{10}) can be determined and used to summarize the system resolution under the given acquisition/reconstruction conditions. It is also important to report the contrast and noise values at which the TTF was measured. This information is necessary not only in terms of the level of CNR needed for a robust TTF measurement (i.e., >15, noted above), but also with respect to the fact that in nonlinear CT systems, TTF can be a function of contrast and noise. Note that z-axis TTF and the slice sensitivity profile[26] reflect similar performance attributes of a CT system. This analysis is illustrated in Figure 6.

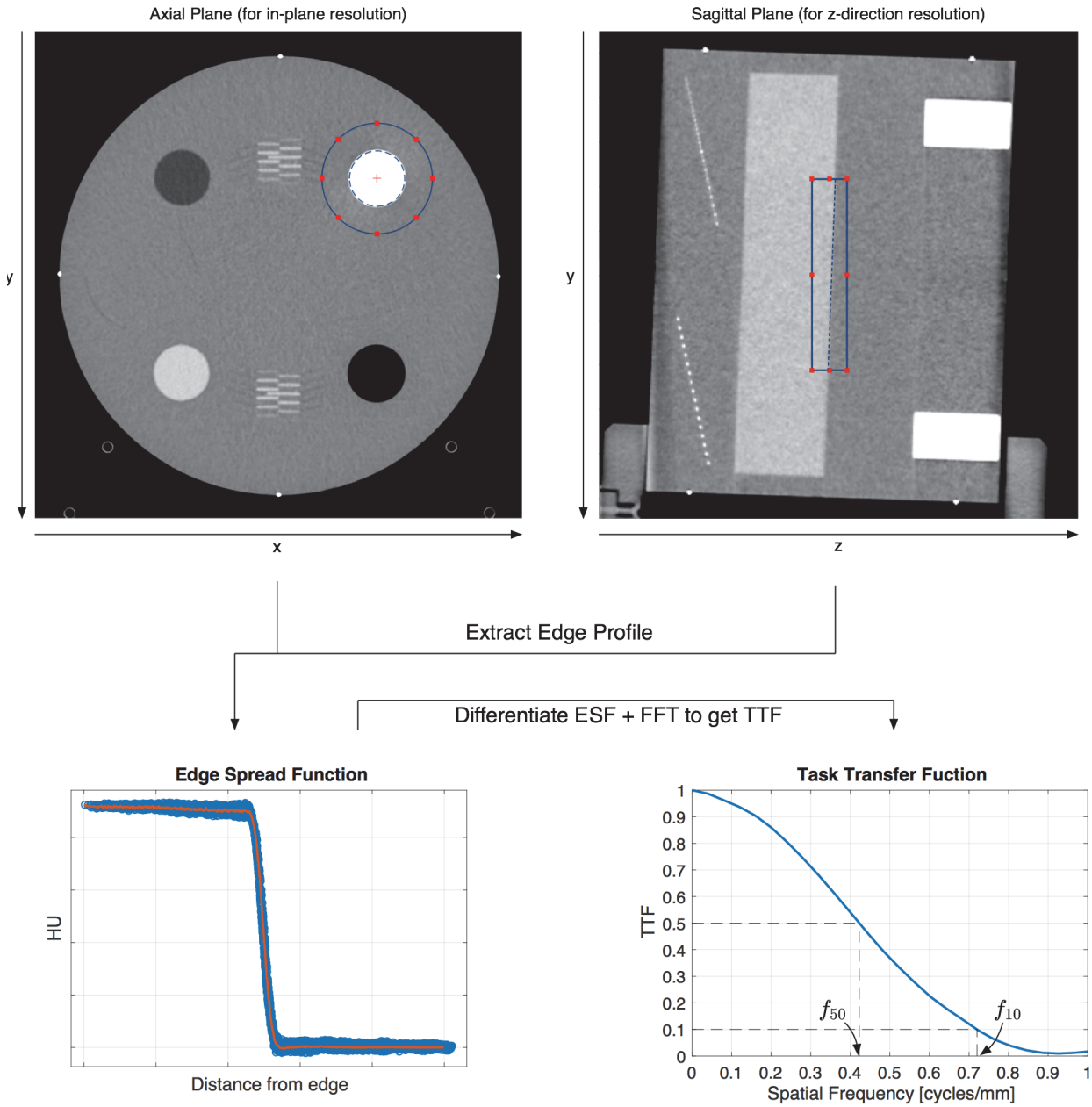

Figure 6. The technique for estimating both the in-plane and z-direction TTF from the CT ACR 464 phantom. The in-plane TTF is measured based on a circular ROI around one of the rods in module 1 (top left). From this ROI, it is possible to identify the center of the rod, and then calculate the distance of each pixel in the ROI from the center. The plot of HU values vs. distance make up the edge spread function (ESF) (bottom left). The data points in the raw noisy ESF (blue dots) are binned and averaged to achieve a smooth ESF (red line). The derivative of the smooth ESF is estimated to get a line spread function, which is then Fourier transformed to get the task transfer function (bottom right). In the z-direction, a cylindrical volume of interest (VOI) is placed around the interface between modules 2 and 3 of the ACR phantom (top right). In this measurement, the phantom is set up with a slight angle relative to the tomographic axial plane. Using voxels within this ROI, the exact location of the edge interface is determined by fitting a plane. It is then possible to extract an ESF by calculating the distance of each voxel from this plane. Using that z-direction ESF, the z-direction TTF is then computed in identical fashion to the in-plane TTF. The TTF curves can be summarized by the spatial frequencies at which the TTF reaches 50% and 10%, denoted as f_{50} and f_{10}, respectively. Various free software resources are available to assist with these analyses (e.g., imQuest, section 5.3).

3.2.6 Precautions and Caveats

When using this method, if CNR_T is below the recommended threshold of 15, then the ESF should be further conditioned to minimize the influence of noise on the TTF measurement using the method described by Maidment et al.[27] Unfortunately, this data conditioning technique assumes that the ESF is monotonic. This assumption is usually violated for reconstruction kernels or algorithms with edge-enhancement and, therefore, this ESF conditioning technique should not be used for images with known or suspected edge-enhancement. Further, CT resolution can be location dependent (shift-variant) and, thus, the results should be ascribed to the radial location where the test object is located. The estimation should include a measure of uncertainty associated with the measurement. Estimation of the uncertainty can be done empirically by making repeated measurements on independent images, or approximated from expected percent error of the TTF as a function of CNR.[24]

Alternative methods for MTF characterization may also use high-contrast wires, beads, or foils (usually made of tungsten).[28] Such techniques, while established for conventional MTF measurements, are not recommended for ascribing the edge properties of low-contrast features when using CT systems deploying nonlinear reconstruction techniques.

3.2.7 Recommended Performance Metrics

- TTF at defined noise and contrast levels in the in-plane and z directions, $TTF_{n,C}$, and $zTTF_{n,C}$. The number of contrast levels (i.e., phantom inserts) needed depends on what the resolution measurements will be used for. For general system characterization, the four CT ACR 464 phantom inserts are sufficient. However, a focused assessment of the resolution properties of low-contrast signals might require a phantom with multiple rods of varying low contrast.
- Frequencies associated with 50% and 10% on in-plane TTF, f_{50} *and* f_{10}, respectively
- Frequencies associated with 50% and 10% on z-axis TTF, zf_{50} *and* zf_{10}, respectively

3.2.8 References

- Chen et al. 2014[24]
- Maidment et al. 2003[27]
- Richard et al. 2012[25]
- Solomon et al. 2015[22]
- Yu et al. 2015[29]

3.3. Noise

3.3.1 Objective

To characterize noise and noise texture of the CT system under reference conditions and establish baseline values for targeted imaging conditions with methods applicable to both linear and nonlinear reconstruction algorithms.

3.3.2 Important Definitions

- Noise: stochastic fluctuations of image pixel values due to measurement uncertainty (i.e., quantum or electronic noise). Fluctuations due to anatomical variations (i.e., anatomical noise) are not considered as noise in this report.
- Noise magnitude: standard deviation (SD) of pixel values
- Noise texture: visual impression of the image noise (e.g., fine or coarse). CT images have a distinct noise texture as a result of the noise correlations introduced in the image reconstruction

process. The noise texture has a large impact on the perceived quality of a CT image and can be quantitatively characterized by the noise autocorrelation or the noise power spectrum as described below.

- Noise autocorrelation: second-order statistic of the noise describing the correlations between any two noisy pixel values. When the noise is wide-sense stationary, the expected (i.e., average) correlation between any two pixels with a given spatial separation is the same, regardless of their absolute location. Wide-sense stationarity of the noise also implies that the noise magnitude is constant across the image FOV. Generally speaking, this condition is not globally satisfied in CT images, but it can usually be assumed to be true within a small local ROI.
- Noise Power Spectrum (NPS): Fourier transform of the noise autocorrelation, describing the distribution of noise variance in terms of spatial frequencies. Noise stationarity is assumed for computation of the NPS from the noise autocorrelation.
- Noise nonuniformity: Variations in the noise magnitude or NPS across the image FOV. All CT images have some degree of global, slowly changing, noise nonuniformity[30] (e.g., noise in the mediastinum tends to be higher than noise in the lungs or the NPS has a different shape at isocenter compared to peripherally). Additionally, nonlinear reconstruction algorithms can introduce highly localized noise nonuniformity in image regions containing many anatomical structures and edges (e.g., lungs). This is due to the regularization used as an integral component of most commercial iterative reconstruction algorithms. As a general rule, these algorithms are attempting to minimize noise while preserving resolution. As a result, they tend to aggressively reduce noise in uniform image regions while less aggressively reducing noise in regions with many structures and edges. This can lead to highly nonuniform spatial distribution of noise.[31]

3.3.3 Equipment

- Water phantom(s) or other uniform phantoms of relevant diameter to mimic the attenuation of a patient's body or head, such as the CT ACR 464 phantom, one of the Catphan phantoms, or the Mercury phantom[23].
- Phantom with "anatomical" texture and structures (i.e., heterogeneous background). A phantom with detailed anthropomorphic structures is preferred (e.g., the Lungman phantom from Kyoto Kagaku, the Mercury texture inserts). If such a phantom is not available, a phantom filled with water and acrylic spheres or other round objects (with different attenuation than water) can be used.

3.3.4 Procedures

3.3.4.1 Noise magnitude

The phantom(s) should be scanned sampling representative protocol and dose conditions of interest or those listed in Table 4. The protocols should ideally range from clinically relevant low dose to typical dose to high dose, as characterization at multiple dose levels enables interpolation of results for the intermediary values to facilitate comparison of results across systems, time, etc. Five ROIs, approximately 1% of the phantom area in size, should be placed at center, and at 12, 3, 6, and 9 o'clock. The peripheral ROIs should be placed approximately one ROI diameter away from the phantom border (see Figure 7). The noise magnitude (SD of pixel values) should be recorded and averaged across each ROI location and for at least three images. This process can use different slices from the same acquisition or the same slice from repeated acquisitions.

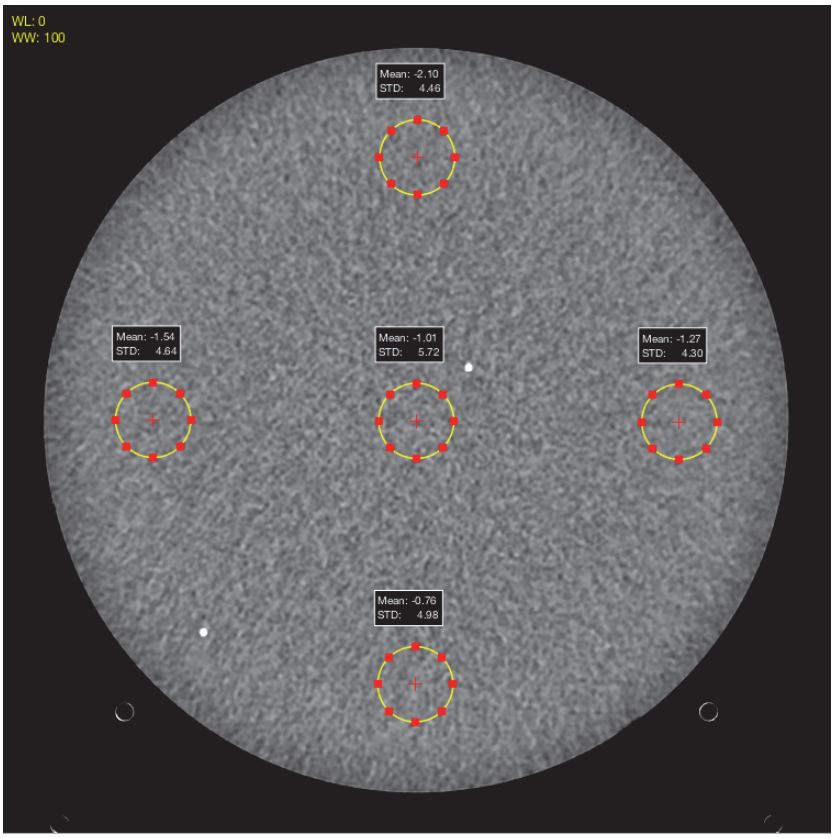

Figure 7. ROI placement for measuring noise magnitude.

3.3.4.2 Noise texture

The phantom(s) should be scanned multiple times using three variations of the typical head and body protocols, including a clinically relevant low dose to a typical dose to a high dose. The number of repeated acquisitions needed depends on the length of the phantom. The target number of ROIs should be 100. A matrix of 64×64 pixels should be extracted near the center of each of the images (being careful to avoid potential artifacts). Multiple ROIs can be used from a single image if needed. The 2D NPS can be estimated from each ROI using the method described by Boedeker et al.[32] The 2D NPS data can be radially re-binned/averaged for 1D presentation.[33] The analysis can also extend to 3D.

3.3.4.3 Noise nonuniformity

A uniform or structured phantom is scanned a minimum of 20 times using a typical head or body protocol in the axial mode at a typical dose[31,34] (the scalar statistic computed depends on the phantom type, see section 3.3.5.2). The scan should be performed with a single rotation and no table translation to minimize variability due to table motion between repeated scans. The spatial distribution of noise magnitude can be estimated on a voxel-by-voxel basis by taking the standard deviation of each voxel's CT number across the ensemble of repeated images as

$$\sigma_i = \sqrt{\frac{1}{M-1}\sum_j^M \left[I_{i,j} - \bar{I}_i\right]^2}, \qquad \text{(Eq. 3)}$$

where σ_i is the noise magnitude of the ith voxel, M is the number of repeated scans, $I_{i,j}$ is the CT number of the ith voxel in the jth repeated image, and $\bar{I}_i$ is the average CT number of the ith voxel across the ensemble of repeated images. σ_i can be thought of as a spatial map of noise magnitude (i.e., how much does each pixel randomly fluctuate from scan to scan). Having this noise map allows one to visually assess noise nonuniformities. Scalar statistics can also be calculated from this noise map as described in section 3.3.5.2.

3.3.5 Data Analysis

3.3.5.1 Noise magnitude and texture

For noise magnitude, the average pixel SD across a minimum of three images is used as the output. The three images could come from repeated acquisitions or from slices (ideally non-consecutive) from the same acquisition. For noise texture, the peak frequency, f_P, and the average frequency, f_A, are reported as summary metrics, which describe the overall frequency content of the NPS as

$$f_P = \arg\max[NPS(f)] \qquad \text{(Eq. 4)}$$

and

$$f_A = \frac{\int f \times NPS(f)df}{\int NPS(f)df}, \qquad \text{(Eq. 5)}$$

where f is the radial spatial frequency (i.e., $f = \sqrt{f_x^2 + f_y^2}$) and $NPS(f)$ is the radially re-binned/averaged 1D NPS (see Figure 8).[33]

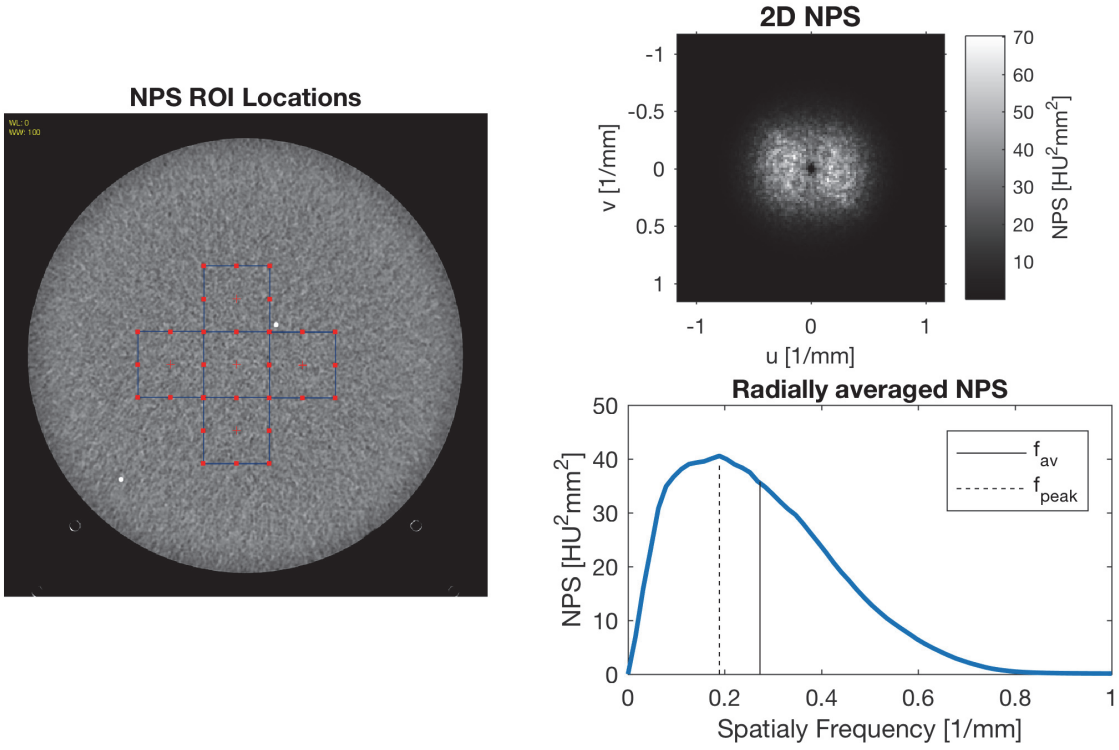

Figure 8. Noise texture analysis using the NPS.

3.3.5.2 Noise nonuniformity

Two scalar statistics have been described in the literature to quantify noise nonuniformity. Both are based on the spatial noise map, σ_i, (see section 3.3.4.3). The first metric was introduced by Li et al. and is called the "noise spatial nonuniformity index (NUI)."[35] This metric is suitable to characterize how much the noise magnitude varies globally across the image FOV. It can be computed based on images of either a uniform or structured phantom. NUI is computed by making a series of ROI measurements (8×8 mm ROI size) spanning the FOV (but within the phantom) on σ_i. For the kth ROI location, the mean noise magnitude, $\bar{\sigma}_k$, is computed, and the NUI is computed as the standard deviation of $\bar{\sigma}_k$ across all ROI locations (i.e., across all instances of k).

$$NUI = \frac{1}{K}\sum_{k=1}^{K}\bar{\sigma}_k, \qquad \text{(Eq. 6)}$$

The second scalar statistic that can be computed is the noise inhomogeneity index, η, as described by Solomon et al.[36] This metric is used to characterize highly irregular/structured spatial distributions of noise, which can appear in iteratively reconstructed images of structured (i.e., nonuniform background) phantoms.[31,36] As such, this metric can only be computed using noise maps from a structured phantom. First, a histogram of σ_i is generated. This histogram should be based on voxels within the phantom only. If the histogram has two distinct peaks, the noise inhomogeneity index, η, is calculated as the relative peak separation divided by the relative height difference in the peaks. Consider the locations of these two peaks on the histogram (x_1, y_1) and (x_2, y_2) is calculated as

$$\eta = \frac{|x_1 - x_2|(y_1 + y_2)}{|y_1 - y_2|(x_1 + x_2)}. \qquad \text{(Eq. 7)}$$

If the histogram does not have two peaks, then η is not defined. Figure 9 illustrates this noise analysis.

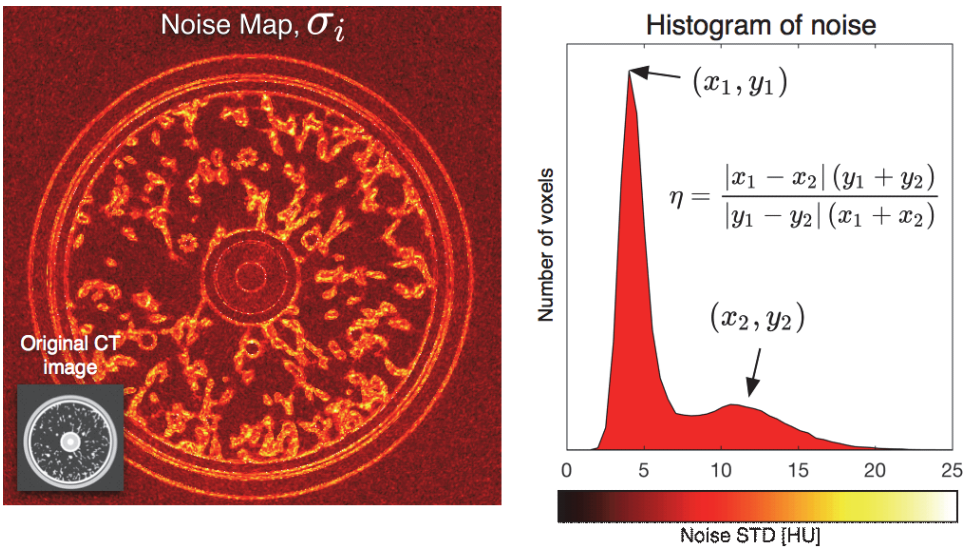

Figure 9. Example of how the noise inhomogeneity index, η, is defined. This metric is meant to quantify the highly irregular and structured spatial distribution of noise that is sometimes observed in iteratively reconstructed images. In this example, a structured phantom simulating lung texture was scanned and images reconstructed using a commercially available iterative reconstruction algorithm. To measure, first a noise map, σ_i, is generated using repeated images of the same structured phantom (see section 3.3.4.3). This noise map gives the standard deviation of the noise on a voxel-by-voxel basis and is shown on the left in this example. Next a histogram of σ_i is generated using only voxels within the phantom, as shown on the right. If the histogram has two distinct peaks, as is the case in this example, then η is calculated using the peak locations and their heights based on the equation shown above. This equation is simply the relative separation divided by the relative height difference between the two peaks. If the histogram does not have two peaks, η is not defined.

3.3.6 Precautions and Caveats

The formulation of the NPS assumes that the noise is wide-sense stationary within the ROI. This assumption may not be valid for very large ROIs due to the known global nonuniformity of CT noise. Ensure the areas sampled for noise determination represent predefined local areas void of artifacts and noise disparity. In the assessment of noise nonuniformity, noise is estimated from an "ensemble" of repeated images. As such, any differences between repeated scans (e.g., phantom motion or different scan settings) will translate into increased variance across this ensemble of images. This could positively bias the measured noise compared to true noise, especially for structured phantoms. Care should be taken to perform the repeated scans in the most reproducible manner possible.

3.3.7 Recommended Performance Metrics

- Noise magnitude (pixel standard deviation) at three dose levels, n
- NPS at defined noise levels, NPS_n
- Peak and average frequencies of the NPS, f_P and f_A
- Noise nonuniformity index, NUI
- Noise inhomogeneity index, η

3.3.8 References

- Boedeker et al. 2007[32].

- Siewerdsen et al. 2002[37]
- Solomon et al. 2012[33]
- Chen et al. 2014[24]
- Li et al. 2014[35]
- Solomon et al. 2014[31]

3.4. Quasi-linear Task-based Performance

3.4.1 Objective

To characterize CT system performance in terms of a Fourier domain-task-based detectability index using the quasi-linear assumption of linear and wide-sense stationary system behavior within a local spatial, contrast, and noise domain. The underlying idea behind task-based image quality assessment is to quantify image quality by estimating how well a human or mathematical observer (i.e., reader) could perform some predefined task (e.g., detection of a subtle signal) on the images in question.[38] Thus a task-based image quality metric is more related to how well an image *performs* in delivering diagnostic information. As a result, task-based image quality metrics are well suited to characterize and/or compare image quality between imaging conditions in which noise magnitude, noise texture, and/or resolution might be variable. CNR on the other hand is only useful as a very simple first-order approximation of low-contrast detectability under fixed noise texture and resolution conditions. In other words, it would not be appropriate to compare different reconstruction kernels or algorithms on the basis of CNR. Specifically, using CNR as the basis of estimating dose reduction potential for iterative reconstruction algorithms (compared to FBP) or different kernels should not be done and could provide highly misleading and suboptimal results.[39] As such, the task-based image quality metrics described in sections 3.4 and 3.5 effectively supersede traditional metrics such as CNR for the assessment of low-contrast detectability in modern CT systems.

3.4.2 Important Definitions

- *Detectability index (d'):* a task-based detection performance metric (often referred to as d' or d-prime). Any task-based image quality assessment technique has three primary components: (1) a task to be performed (usually the detection of a subtle lesion/signal), (2) an observer to perform the task (typically a mathematical detection algorithm, sometimes a human reader), and (3) images to be assessed. Based on the foundational mathematics of signal detection theory,[38] one can imagine an ensemble of images acquired under identical conditions, some with a target signal to be detected and some without. The observer processes the image data and outputs a scalar response variable for each image, proportional to the observer's confidence that a signal is present. This results in two distributions of the response variable (signal-present and signal-absent). The greater the separation between those distributions, the better the observer is at correctly detecting the signal. The detectability index, d', quantifies this degree of separation and is essentially the signal-to-noise ratio of the observer's response variable for the aforementioned signal-present and signal-absent distributions. Mathematically, the square of d' is the squared difference in the distribution means, divided by their average variances. One would expect a different d' for different tasks, different observers, and/or images acquired under different conditions. In this section, the task considered is the detection of a circular signal as defined by a task function (W_{task}, see below) and the mathematical observer used is a non-prewhitening matched filter (NPW, see below). It turns out that one can compute the detectability index, d'_{NPW}, for this observer in the Fourier domain based on measurements of system resolution (TTF), and noise (NPS) as shown below. To do this calculation, it is also necessary to define the properties (e.g., size, shape, contrast, and contrast-profile) of the signal be detected. These properties are

encoded in the task function, W_{task} (see below). The detectability index is interpreted as a metric of image quality due to its relation to low-contrast detectability (i.e., an increase in d'_{NPW} implies the signal is easier to detect and thus image quality is better). Such Fourier-domain metrics have been shown to agree closely with human observer studies for linear shift-invariant (LSI) systems[16] and are being adapted for use in nonlinear systems that are evaluated in a quasi-linear state (e.g., IR algorithms).[19,22,40]

- *Task Function (W_{task}):* Fourier transform of the signal to be detected (e.g. a 10-mm circular lesion with a contrast of 10 HU). As mentioned above and described in detail below, measuring d' involves defining the properties (size, shape, contrast, and contrast-profile) of the signal to be detected. These properties can be encoded in this task function. Common task functions correspond to circular low-contrast signals having diameters between 1 to 10 mm. Suggested mathematical formulations are provided below.
- *Non-prewhitening matched filter (NPW):* this observer model compares the image of interest to a template consisting of the expected signal via cross-correlation. This model has been shown to correlate strongly with human performance for low-contrast detection tasks.[19,22,40] The detectability index for this model, d'_{NPW}, can be computed in the Fourier domain for a given W_{task} based on measurements of the system's TTF and NPS as shown below.
- *Estimability index (e')*: a performance metric related to the expected accuracy of volumetric measurements in CT images with given three-dimensional noise and resolution properties. A higher e' implies a higher degree of expected volumetric accuracy.

3.4.3 Equipment

- Phantom with circular inserts (>2 cm diameter) of various materials (i.e., contrast levels). Phantom should also contain uniform regions for noise analysis (e.g., CT ACR 464 phantom for constant size or Mercury phantom for variable size measurements).
- Image analysis software capable of NPS, MTF/TTF, and detectability calculations, see section 5.3

3.4.4 Procedures

Position and align the phantom on the table. Acquire a CT localizer radiograph and define the scan range to incorporate the entire phantom. The phantom is scanned following protocols interest representative of the protocol and dose conditions and/or those listed in Table 4. The protocols should ideally range from clinically relevant low dose to typical dose to high dose.

3.4.5 Data Analysis

For each acquisition and phantom insert, estimate the TTF using the methods described in section 3.2. Define a task function, W_{task} by first synthesizing an ideal image of a signal to be detected. Then W_{task} *is defined as the Fourier transform of this synthesized signal.* As common approximations, the shape of the signal may be circular with either a rectangular or designer contrast-profile (Figure 10),[24,41] respectively, formulated as

$$c(r) = C\prod(rD/2), \text{ or } c(r) = C\left[1-\left(\frac{2r}{D}\right)^2\right]^n, \qquad \text{(Eq. 8)}$$

where c is the value of the signal at a radial distance r, C denotes the peak contrast of the signal against the background in HU, $\prod$ is the rect function, D is the signal diameter, and n is a constant dic-

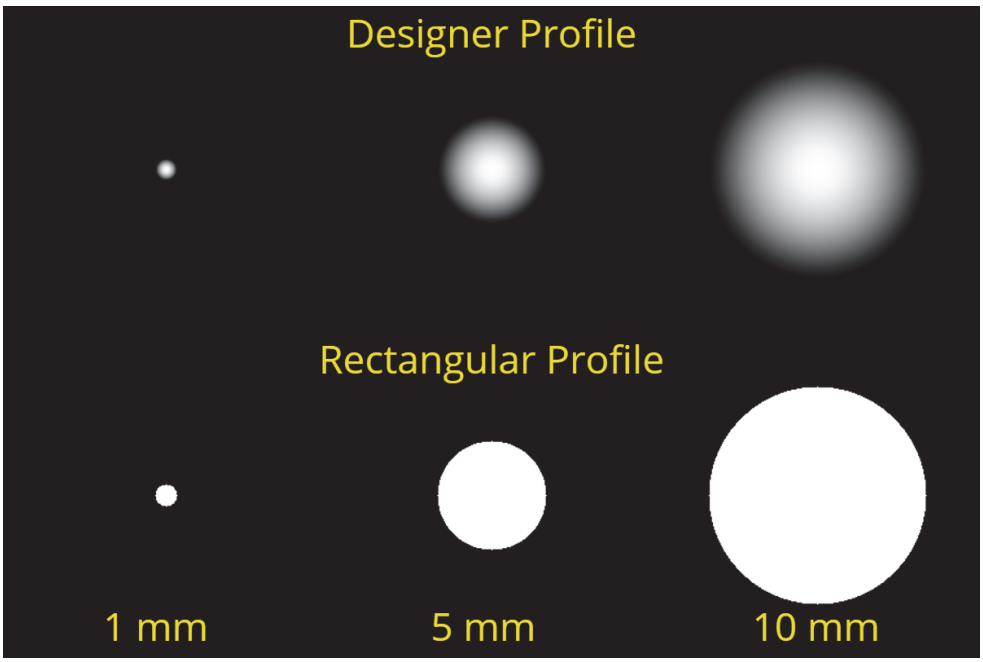

Figure 10. Examples of the synthesized signals to be detected. The Fourier transform of such a signal is the task function, W_{task}, which is an important component of the detectability index calculation. Signals of three sizes are shown, with the top rows showing signals with a designer contrast profile and the bottom row with a rectangular contrast profile.

tating the sharpness of the edge of the designer contrast-profile (as n decreases, edge sharpness increases, recommended value is 1, with alternative values ranging between 0.25 to 2). Note that in the designer profile equation, D denotes the diameter at which the contrast reaches zero. The apparent (i.e., visual) diameter of the signal might be lower than D depending on the chosen value of n. In some cases, it is useful to parameterize the object's diameter by its full-width-at-half-maximum (FWHM), which can be computed as a function of D:

$$FWHM = D\sqrt{1-0.5^{1/n}}\,. \qquad \text{(Eq. 9)}$$

Three diameters for the signals are recommended: large (10 mm), medium (5 mm), and small (1 mm). Note that the contrast, *C,* of the signal should roughly match the contrast of the insert used to measure the TTF. An ideal range is between 10 and 100 HU. Next calculate W_{task} by taking the Fourier transform of the synthesized image. Finally calculate d'_{NPW} as

$$d'^{2}_{NPW} = \frac{\left[\iint \left|W_{task}(u,v)\right|^2 \times TTF^2(u,v)\,du\,dv\right]^2}{\iint \left|W_{task}(u,v)\right|^2 \times TTF^2(u,v) \times NPS(u,v)\,du\,dv}\,. \qquad \text{(Eq. 10)}$$

where u and v are the spatial frequencies corresponding to the x and y direction, respectively. This formulation of the detectability index, based on the non-prewhitening matched filter, assesses detection in two-dimensional images, thus using 2D integrals. The analysis may be extended to 3D using the 3D TTF and NPS. Using the recommended protocols above, it is possible to assess image quality as a

function of dose, tube potential, tube current modulation setting, phantom size, task size, task contrast, image thickness, reconstruction algorithm, and reconstruction kernel, thus offering a system characterization over a wide sampling of operational settings.

The analysis can further be extended to estimability for specific estimation tasks such as assessment of a lesion volume. For those tasks, an estimability index can be computed as

$$e'^2 = \frac{\left[\iiint \left|W_{task}(u,v,w)\right| \times \left|M(u,v,w)\right| \times TTF^2(u,v,w)\,du\,dv\,dw\right]^2}{\iiint \left|W_{task}(u,v,w)\right| \times \left|M(u,v,w)\right| \times TTF^2(u,v,w)NPS(u,v,w)\,du\,dv\,dw}, \quad \text{(Eq. 11)}$$

where W is Fourier transform of the derivative of a spherical signal's edge profile in the definitions of Eq. 8 (constructed in 3D), and M is a template function associated with a lesion segmentation algorithm.[42] The template function reflects the contribution of the lesion segmentation algorithm to the estimation process (volume estimation in this case). This function is expected to be reflective of the method employed by the specific segmentation algorithm. In this case, a typical algorithm is assumed to be seeking spherical features as a basis of the segmentation.

3.4.6 Precautions and Caveats

When creating W_{task}, be sure that the assumed contrast of the object to be detected and the insert used to measure the TTF are similar. Also be sure to report the contrast of the insert and noise conditions under which the TTF was measured.

Non-prewhitening matched filter is only one among many such observer models that can be used to integrate and extend the resolution and noise properties of an imaging system toward the performance for a defined task[48]. In this report we note the use of only one model (NPW) to standardize the process based on a model that has shown strong correlation with observer data[19,45]. Future extensions may include other models provided the details are disclosed and an efficient analysis strategy is made available.

The Fourier–based methodology to characterize imaging system performance assumes a quasi-linear shift-invariant system response and locally wide-sense stationary noise statistics. Thus, the testing conditions and image and noise features should be carefully selected to closely match the patient imaging conditions.

Using all combinations of settings in Table 4 will result in 96 CT series and multiple d' values depending on the phantom used (384 if using the CT ACR 464 phantom, 2400 if using the Mercury phantom). The interpretation of this complete dataset can be practically prohibitive for routine testing but with automation can be done for the initial system characterization and updated as a basis for comparing or optimizing clinical protocols.

3.4.7 Recommended Performance Metrics

Detectability indices for the task of detecting target reference circular signals (e.g., 1, 5, and 10 mm features at specific CT numbers with rectangular or designer morphology) for the targeted phantom size and noise or dose level.

Estimability index for the task of estimating the volume of a target reference spherical signal (e.g., 10 mm feature at specific CT number with rectangular or designer morphology) for the targeted phantom size and noise or dose level.

3.4.8 References

- Chen and Samei
- Christianson et al. 2015[19]

- Gang et al. 2011[16]
- ICRU Report 54[38]
- Samei et al. 1997[41]
- Solomon et al. 2015[40]
- Solomon et al. 2015[22]

3.5 Spatial Domain Task-based Performance

3.5.1 Objective

To evaluate a CT system in terms of the ability of the images to enable an observer (either a human or a computer algorithm) to perform a signal-detection task using the image pixel values themselves. Spatial domain methods are available for characterizing signal detectability for known signals, as well as detection of signals that have unknown or variable aspects, including size, contrast, or location. Tasks in which the signal isn't exactly known to the observer, because the signal is deterministic but the observer has missing information about it (where the signal is, or its size or amplitude), are more similar to clinical tasks. Even more similar to a clinical task are ones in which the signal is random in some way. Tasks with variable or unknown signal characteristics can offer advantages over signal-known-exactly tasks when evaluating image quality, including the need for fewer images as well as the ability to make use of signals with a larger range of contrast levels while still resulting in meaningful comparisons between image acquisition protocols or reconstruction algorithms.

3.5.2 Important Definitions

- *ROC:* receiver/relative operating characteristic
- *LROC*: localization relative operating characteristic
- *FROC*: Free-response operating characteristic
- *EFROC:* exponential transformed free response operating characteristic
- *MRMC*: multiple-readers multiple cases study
- *AUC*: area under the curve of any of the operating characteristic curves above

3.5.3 Equipment

- *FDA MITA Phantom,* as described in section 5.2.5. The FDA MITA phantom has a set of four low-contrast embedded rods that facilitate the evaluation of CT systems in a single image. Using rods as targets in the phantom also allows multiple images to be utilized in a system evaluation from a single reconstructed 3D image.
- Catphan Custom Modules: A phantom design consisting of five identical spherical signals arranged symmetrically about the center, as shown in section 5.2.5.[43] Modules with signals of different size or contrast can be used depending on the scanner type or operating regime being evaluated. This design has the advantage that it can be used for 3D image evaluations because the spherical targets are limited in the z-axis, unlike the rod objects utilized in the FDA MITA phantom.
- The practitioner may consider using other phantoms containing objects with sizes and contrasts relevant to the task being evaluated. This includes anthropomorphic structured phantoms with embedded signals. Note that the specifics of the analysis method may somewhat depend on the phantom design (e.g., number of repeated images needed, 2D vs. 3D signals, etc.).

3.5.4 Procedures

The phantom is imaged using the representative protocol and dose conditions of interest or those listed in Table 4. The images are scored by human observers or computational algorithms (so-called observer models). The specifics of collecting the scoring data from either method are outlined below.

This characterization mostly applies to situations when a multiplicity of conditions may need to be compared (e.g., effect of dose on signal detection). The number of images required is based on the desired statistical power of the final results; the targeted operating points and the performance difference can be compared between the conditions. As an example, with the custom module arrangement,[43] 20 signal-present and 20 signal-absent image samples were needed for showing the difference in performance between a standard filtered back-projection implementation and an iterative reconstruction algorithm.

3.5.4.1 Human observer methodology

Both the FDA MITA phantom and the Catphan custom modules are designed to allow for unknown-location signal detection experiments with human observers. To set up such an experiment, the evaluator crops regions of interest (ROIs) from the reconstructed images such that a signal is located within the ROI, with a chosen distance from the ROI edge to avoid boundary effects. The ROIs are selected such that each contains only one signal. Furthermore, for a set of image scans, the locations of the boundaries of the ROIs are adjusted so that the signals have different locations within the ROIs, making it so that the signal locations are random to the observer when presented as a sequence of such ROIs. Figure 11 shows such an example of random 5×5 cm^2 ROI selections applied on an FDA MITA phantom image, and cm^2 ROI randomly selected on Catphan custom modules. A different set of such ROI selections can be made for each of a set of image scans, enabling the signal to be located at different positions in the set of ROIs for an observer signal-detection experiment with signal-location uncertainty.

It should be noted that for the FDA MITA phantom, which contains four different signals with different size-contrast values, the detection of each signal type could be studied separately, with ROIs that overlap across the signal size-contrast experiments (Figure 11a). However, if the results for different signals are subsequently combined, for example, to examine the impact of signal size for a given imaging protocol or reconstruction algorithm, then the possible correlations due to overlap between the ROIs must be accounted for in the calculation of the error bars.

Because the signals in a Catphan custom module are identical, and meant to be studied together, the ROIs should not overlap (Figure 11b) in order to assure independence of each ROI reading.

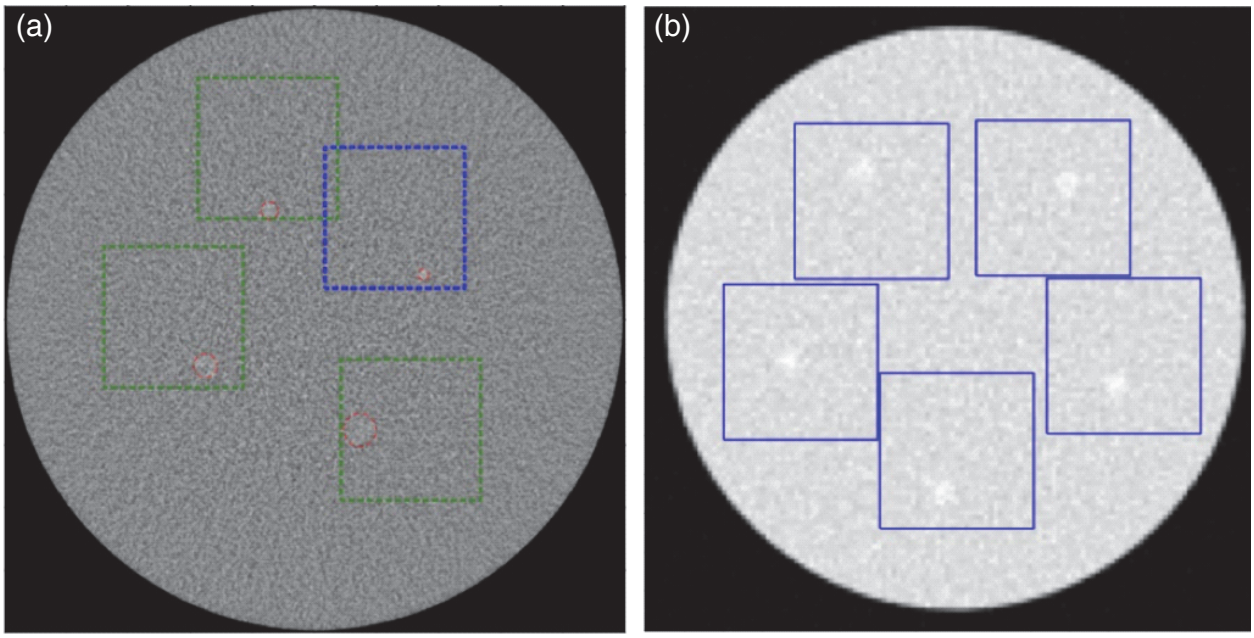

Figure 11. Example of random selection of regions of interest (ROI). (a) 5×5 cm^2 ROIs in FDA MITA phantom; the ROIs are allowed to overlap. (b) 4×4 cm^2 ROIs in a Catphan custom module with identical signals; the ROIs are not allowed to overlap.

Once the ROIs have been specified, the next step is to present them to the human observer for data collection. In an LROC study, the reader views signal-present and signal-absent ROIs individually and is asked to rate the probability that a signal is present and indicate the signal location. (Of course, the reader is blind to whether a signal is truly present in the ROI, and if it is present, where it is located.) A simpler experimental design is to display only the signal-present ROIs and ask the reader to indicate the signal location. In this case, the correct-localization success rate is the performance metric.

The simplest experiment of all is the location-known-exactly experiment, in which the observer is told where the signal would be if it were present, as well as a full description of it size, shape, contrast, etc. The observer simply scores each ROI using a scale that indicates their certainty that the signal is present. While this experimental design is common, the extent of information provided to the observer makes the signal highly detectable for all but very low-contrast situations when the background is also known, as it is in the phantoms described here. It should be noted that the advent of 3D printing is bringing new opportunities for creating phantom images with nonuniform backgrounds, which is showing promise for the use of known-location signal-detection experiments at higher contrasts than are feasible in flat-background phantom experiments.

It is widely known that human observers vary in skill, and thus it is important when evaluating an imaging system with human observers to use a sample of readers so that the study results generalize to other readers and not just the one(s) used in the study. The term “Multi-Case Multi-Reader” (MRMC) study design is used to refer to such studies, where “case” refers to the images used in the evaluation experiment here.

3.5.4.2 Observer model methodology

Mathematical observer models are a powerful alternative to human observers for CT performance evaluation. Section 3.4 described one such example, where the figure of merit was based on an observer model that evaluates the images in the Fourier domain. In the spatial domain, the observer model is one that takes in the image pixel data and applies an algorithm that is typically inspired by human perception experiments. The literature describes spatial-domain observers that have been shown to be very good predictors of human performance.[18,44–49] The task performed by the observer model can involve a signal searching (or image scanning) algorithm. Several approaches are possible for signal searching. One such procedure consists of applying a signal-matching template at all locations of a given image area (or volume). Subsequently, the practitioner can retrieve the list of the most suspicious locations following one of the methods.[43] The results can be analyzed as described below. In the case of the FDA MITA phantom, special area search restrictions need to be applied, so that the search for a given signal size, using a specialized template, would not be confused by the presence of the other signals of different size.

3.5.5 Data Analysis

Whether the observer in a spatial-domain evaluation is a human or an observer model, the data analysis method is virtually identical. Below we describe the data analysis methods for various experimental designs and provide references for more information. Publicly available software resources are noted in section 5.3.

3.5.5.1 Localization success and LROC analysis of ROI reading scores

If only the signal localization marks are recorded, then the localization success rate can be used as a performance metric. More information is acquired in an LROC study with readers providing confidence scores. The data can be analyzed using LROC methods, with the area under the LROC curve (LROC-AUC) as the performance metric.[50,51] An adaptation for LROC of the method[52] can be used for an MRMC study design.

3.5.5.2 Free-response data analysis of the automatic signal-search scores

The results returned by the image scanning procedure follows a free-response image reading methodology: the observer marks and scores all "potentially" suspicious locations (above a certain score or probability of signal presence), and the data can be analyzed using free-response operating characteristic (FROC) analysis. A variation of this method using the exponential transformation of abscissa, EFROC,[53] is particularly advantageous for use with this type of phantom data.

In many circumstances there is a need to evaluate image quality for a multiplicity of conditions, for example, across a broad range of doses to determine the dependence of signal detectability on dose. The number of images required for such an evaluation will be based on the desired statistical power for drawing conclusions regarding statistical significance of the difference between conditions. As an example, with the Catphan custom module arrangement,[43] 20 signal-present and 20 signal-absent image samples were needed for showing the difference in performance between a standard filtered back-projection implementation and an iterative algorithm.

3.5.6 Precautions and Caveats

The current FDA MITA phantom allows only 2D evaluations, and the effective noise level for a given dose depends on image thickness. Ideally, doubling the thickness is equivalent to doubling the dose. However, rebinning, interpolation, filtering (along the z-axis), regularization, and other procedures that are algorithm-dependent may lead to a different *effective* image thickness than the *nominal* image thickness that is defined by the chosen size of the voxels. As a result, the signal detection performance may vary with the image thickness for the compared conditions. In order to assess this effect, we sug-

gest measuring the *z*-axis resolution at low contrast.[24] The z-axis resolution for the algorithms being compared should be shown to be non-inferior to that of the standard algorithm. A recent review paper by Vaishnav et al. consolidated information relevant to objectively assessing iterative algorithms and their dose reduction potential using task-based methodologies.[54]

A general aspect of using signal-matching templates as observer models for search tasks is that they may not necessarily yield optimal detection performance. This is due to the fact that the optimal performance depends not only on matching the signals, but also on avoiding the false-signals, which itself depends on the noise pattern as well as on the template used. As opposed to the case of the location-known detection tasks in which the optimal matching template can often be analytically derived (or estimated), in search-matching templates, the statistical population of false-signal locations depends on the template used and, thus, template optimization is nontrivial, requiring a heuristic approach. As the signal-searching template depends on a number of parameters (e.g., search window size), a generally recommended approach for showing the optimality and stability of the results is to do a sensitivity and uncertainty analysis. It should be noted that such an analysis is necessary for interpretation of the results obtained via any observer model or detection paradigm.

Because signal search tasks take into account the rather extreme occurrences of false, signal-like, features randomly appearing in the image, they are particularly sensitive to changes in the noise magnitude and pattern, in addition to emulating more closely some clinical diagnostic tasks. When applied with phantoms as presented here, the signal search task procedures make better use of the image area, thus being more efficient in terms of assessing image quality with a finite amount of image data.

The number of images needed to calculate spatial domain task-based image quality metrics could limit their utility for routine clinical physics testing, especially when exploring a large parameter space of system acquisition and reconstruction settings.

There is a large body of scientific literature related to the use of observer models as the basis of image quality assessment in medical images. For practical purposes, only a small portion of that literature is represented in this report. Interested readers are encouraged to review ICRU Report 54[38] and a comprehensive review article by Barret et al.[48] for a more in-depth understanding of task-based image quality assessment and different types of observer models.

3.5.6.1 Fourier and spatial-domain observers

Fourier and spatial-domain observers strive to predict task-based measures of image quality, while accounting for characteristics of the signal to be detected as well as the deterministic and stochastic properties of the imaging system. Spatial-domain methods are applied directly to the acquired images. Publicly available software allows for the calculation of performance estimates and their statistical uncertainties (see iMRMC and iQModelo in section 5.3). The number of images needed for reasonable error bars supporting statistical comparisons between conditions will depend on the inherent detectability of the signal, the number of ROIs per image, and the size of the search area in the case of a search task.

The Fourier methods described in this report make use of measures of the characteristics of the imaging system in terms of resolution and noise, and a subsequent calculation to derive a task-based figure of merit. The calculation of the uncertainty of the resulting figure of merit requires an understanding of the uncertainties of each of the underlying components of the performance estimate, and the impact of that uncertainty on the final estimate.

Ongoing work in the field is addressing the need for more realistic (virtual and physical) phantoms that more closely resemble clinical tasks.[31,34,55–60] The literature regarding the use of observer models that correlate well with human performance continues to grow and mature, elucidating the circumstances for which Fourier or various spatial-domain observer models are useful surrogates of human performance. Finally, a more complete understanding of the relationship between the numbers of

scans used for each approach and the resulting uncertainty in the performance metric is an area of active investigation.[24,61]

3.5.7 Recommended Performance Metrics

- Localization success rate for targeted tasks, *LR*
- Area under the LROC curve for targeted tasks, A_{LROC}
- Area under the EFROC curve for targeted tasks, A_{EFROC}

3.5.8 References

- Gallas 2006[52]
- Popescu 2007[50]
- Popescu 2011[53]
- Popescu and Myers 2013[43]
- Wunderlich and Noo 2012[51]
- Vaishnav et al. 2014[54]
- Tseng et al. 2014[62]
- Yu et al. 2013[47]
- Leng et al. 2013[18]

4. Clinical Utility and Future Extensions

The characterization procedures and resulting metrics delineated in section 3 act as a first step in providing common analysis techniques and metrics that can characterize the anticipated clinical performance of a CT system based on physical measurements. The tests described in section 3 provide advanced methods to assess modern CT systems that have advanced technologies such as AEC and iterative reconstruction available. As noted earlier, the goal of these procedures are not so much to verify the technical or engineering specifications of CT systems, or to pass or fail a device, but rather to characterize the system in terms that aid in improving and optimizing its clinical utilization. That is the rationale, for example, to ascertain the system performance not just in terms of tube potential accuracy, but in terms of resolution or task-based detection.

The utility of the operational performance noted above can be exemplified in a few specific applications.

It is not uncommon for providers of radiological services to have a diverse fleet of CT equipment with scanners of different makes and models. This diversity poses a challenge in providing images with consistent quality due to differences between scanner models. For example, images from different scanner models tend to have a unique visual impression. This difference in the overall "look" of the images is especially noticeable across different CT manufacturers and is due in large part to differences in image reconstruction methods that then lead to differences in noise texture. The metrology delineated in this report can be used as a basis to match protocols across CT systems of different makes and models, providing a method to adjust the acquisition parameters for each system such that the systems would deliver consistent image quality within a target range.

Because noise texture is one of the primary image properties that defines a human reader's visual impression of the image, being able to achieve similar noise texture across scanners models helps to achieve much more consistent images. The primary factor that determines the noise texture is the kernel used in image reconstruction. Kernels could be matched across scanners from different manufac-

turers to achieve similar noise texture based on noise power spectrum (NPS) analysis similar to that described in this report.[33] This matching can be expanded to include resolution (TTF) and noise magnitude as well, while at the same time meeting constraints in the dose level of the examination.[63] This method relies on the f_A and f_{50} metrics of noise and resolution that are described in sections 3.2 and 3.3 of this report, respectively. An illustration of what this might look like in practice is shown in Figure 12.

One step beyond achieving *consistent* image quality is the goal of *optimizing* image quality. The metrics derived from the tests in this report could be used to ascertain the system performance as a function of protocol parameters, so that a desired image quality and dose can be targeted based on the indication and the patient attributes for existing or new protocol definitions. Fourier-domain task-based image quality metrics similar to those described in section 3.4 of this report could be used as a tool to help balance the competing demands of image quality and radiation dose when defining CT protocols for a large multivendor clinical facility.[64] Estimating detectability for a large variety of clin-

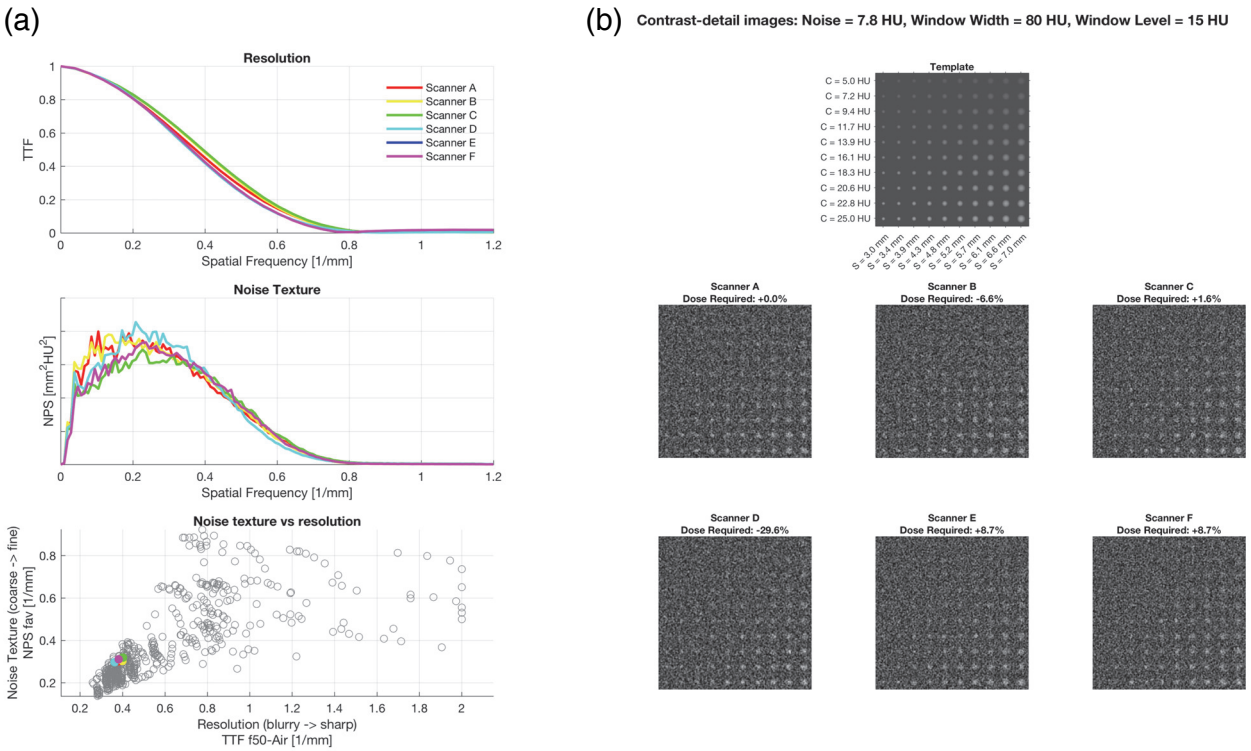

Figure 12. An example illustrating how the advanced metrics described in this report could be used in practice to help match image quality across different scanner models. The left panel (a) shows the TTF (top), NPS (middle), and f_A vs. f_{50} (i.e., noise texture vs. resolution) for six different scanner models (A-F). Each scanner's reconstruction settings were chosen to achieve similar resolution and noise texture properties based on minimizing differences between f_A and f_{50} across scanner models. The gray circles in the scatter plot each represent a possible reconstruction kernel or iterative setting available on the scanners, and the colored dots represent the chosen reconstructions corresponding to each scanner. The right panel (b) shows a series of synthesized contrast-detail images that represent what images from each scanner would look like given their TTF and NPS, and with dose adjusted to achieve equal noise magnitude as that of scanner A at a reference dose. The contrast-detail diagrams let one visually assess the smallest and lowest contrast signal that one would expect to be able to detect given the noise and resolution properties of each scanner. Visual inspection of these images confirms similar expected detectability across scanner models.

ical protocols, patient sizes, lesion sizes, and lesion contrasts, the dose needed to detect a lesion of a given size and contrast for a given patient size can be ascertained and displayed to help a user estimate how changes in radiation dose would be expected to affect detectability of targeted lesions across a patient population (Figure 13a). Similarly, metrology can be applied using spatial-domain task-based image quality metrics to determine the achievable dose reduction of differing reconstruction algorithms.[49] Scanning a phantom with low-contrast signals at various dose levels, reconstructing the data with FBP and iterative algorithms, the magnitude of possible dose reduction using iterative reconstructions can be ascertained based on desired detectability (Figure 13b).

Performance evaluation as detailed in this report advances the field beyond specifications toward actual clinical performance. It paves the way toward a number of possibilities, even beyond the few

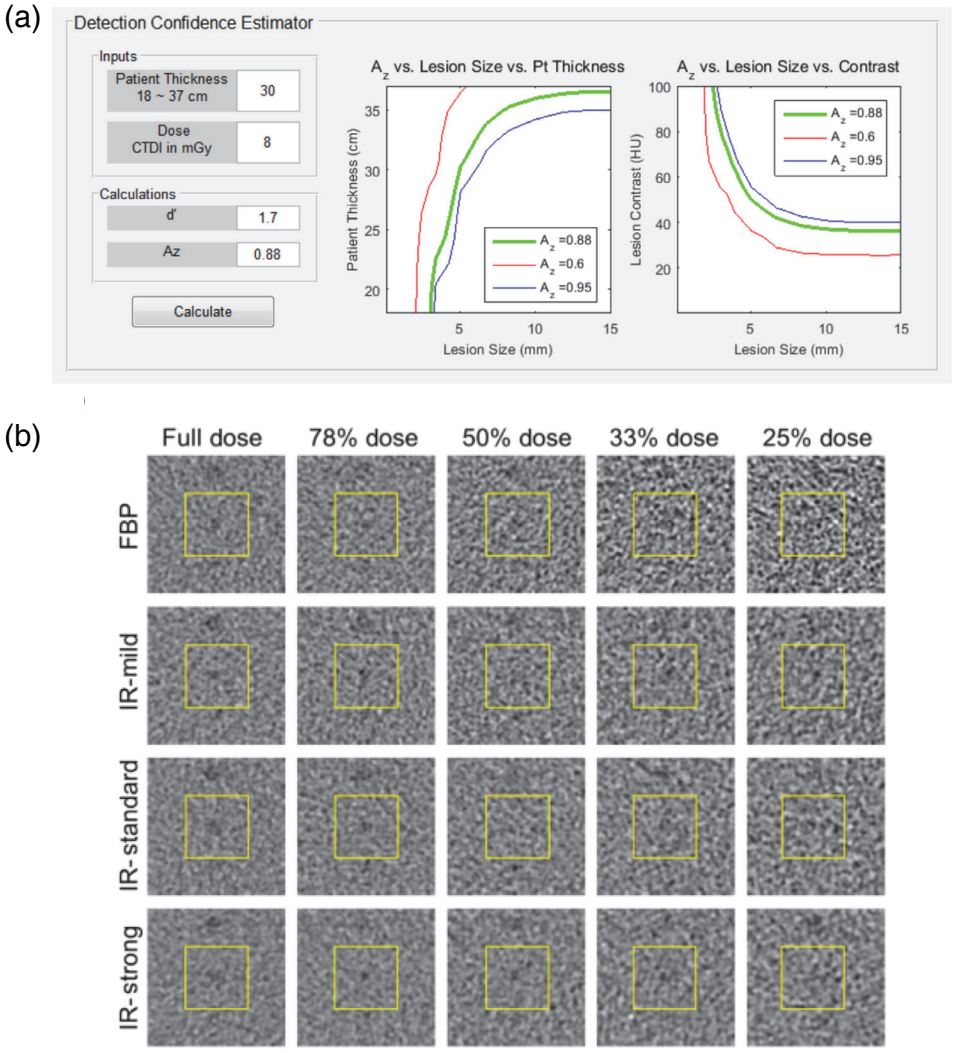

Figure 13. Example of (a) the graphical interface developed by Zhang et al.[64] to help a user balance the trade-offs between radiation dose and image quality (i.e., detectability) as a function of patient size and lesion characteristics. Also an example of (b) the trade-offs between radiation dose, reconstruction method, and detectability as demonstrated by Favazza et al.[49] That work determined the detectability of low-contrast signals under all the shown conditions using spatial-domain task-based image quality metrics. Those detectability data were then used to determine the achievable dose reduction from the iterative reconstruction algorithm in question.

noted above. These possibilities have not been fully delineated within the present report, but they offer exciting prospects for the future direction of CT metrology and clinical optimization:

1. *Performance monitoring*: The methodology provides the means to monitor the performance of a CT system in terms of metrics that are more directly related to clinical performance, as opposed to engineering attributes.
2. *Quantitation*: Task-based performance evaluation lends itself to defining tasks beyond detection, thus enabling the use and conformance of CT systems to provide precise quantitative output. Characterization of CT performance for the task of estimating lesion volume was briefly discussed in this report, but more work is required to extend that characterization to other important areas of quantitative CT imaging, such as measuring the shape, texture, or material composition of a lesion.
3. *Benchmarking*: The methodology enables benchmarking the performance of a CT imaging operation in terms of its similarity to other practices (e.g., in peer institutions) or desired clinical performance.
4. *Registries*: The methodology provides quantitative values to be used for CT image quality registries within state, national, or international systems.
5. *System development*: In design and construction of CT systems, the system can be designed and calibrated based on targeted image quality output beyond engineering specifications, the dependence of which to clinical performance is less certain.
6. *Conformance*: The metrics can be used as a basis for accreditation and conformance validation of CT operations to desired targets to improve consistency across CT operations.

The above goals can be approached by conducting data-collection trials from existing imaging operations using the methods described in section 3, so that target performance values can be ascertained (e.g., the desired f_{50} for a clinical CT system, protocol, or indication). The future direction of the material presented in this report falls along this pathway.

5. Supplemental Information

5.1 List of Acronyms

AAPM	American Association of Physicists in Medicine
ACR	American College of Radiology
AEC	Automatic Exposure Control
AUC	Area under the Receiver Operating Characteristic (ROC) curve
CFR	Code of Federal Regulations
CMS	Centers for Medicare & Medicaid Services
CNR	Contrast-to-Noise Ratio
COV	Coefficient of Variation
CR	Computed Radiography
CT	Computed Tomography
CTAP	Computed Tomography Accreditation Program
CTDI	Computed Tomography Dose Index
DICOM	Digital Imaging and Communications in Medicine
DLP	Dose Length Product
EC	European Commission
EFROC	Exponential-transformed Free Response Operating Characteristic.

ESF	Edge Spread Function
FBP	Filtered Back Projection
f_{10}	Spatial frequency associated with 10% power of the MTF or TTF
f_{50}	Spatial frequency associated with 50% power of the MTF or TTF
f_A	Average spatial frequency of the NPS
f_p	Peak spatial frequency of the NPS
FDA	Food and Drug Administration
FFT	Fast Fourier Transform
FOV	Field of View
FROC	Free Response Receiver Operating Characteristic
FWHM	Full Width at Half Maximum
HVL	Half-Value Layer
HU	Hounsfield Unit
IAEA	International Atomic Energy Agency
ICRU	International Commission on Radiation Units & Measurements
IEC	International Electrotechnical Commission
IR	Iterative Reconstruction
LSF	Line Spread Function
LROC	Localized Receiver Operating Characteristic
LSI	Linear Shift-Invariant
kV	Tube potential, in kilovolts, across x-ray tube
LSI	Linear Shift Invariant
mA	Tube current, in units of milliamperes
mAs	Tube current-time product, in units of milliamperes-seconds
MDCT	Multiple detector computed tomography
MITA	Medical Imaging Technology Alliance
MRMC	Multiple-readers multiple case ROC metrology
MTF	Modulation Transfer Function
N	actual number of data channels used during one axial acquisition
NEMA	National Electrical Manufacturers Association
NPS	Noise Power Spectrum
NPW	Non-prewhitening match filter
OSL	Optically stimulated luminescence
PACS	Picture Archiving and Communication System
QA	Quality Assurance
QC	Quality Control
QMP	Qualified Medical Physicist
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SD	Standard Deviation
SSP	Slice Sensitivity Profile
T	width of each channel ($N \times T$ = nominal radiation beam width)
TCM	Tube Current Modulation
TG	Task Group
TTF	Task Transfer Function
WL	Window Level
WW	Window Width

5.2 Phantom Examples

This section provides some examples of phantoms that could be used for the analyses described in this report. It is not intended to be an exhaustive list of all available CT phantoms, nor is it an official endorsement by the AAPM of any of the phantoms mentioned below. Rather, the goal of this section is to provide several examples of phantoms to demonstrate the properties and features that a phantom would need in order to perform the tests described in this report. For each phantom, a short description is given, along with a list of tests that could be performed with that phantom and some representative photos and CT images. For the phantoms that are commercially available, detailed specifications can be obtained from the manufacturers and are not given below; specific model numbers and manufacturers are listed at the end of this section.

5.2.1 ACR CT Accreditation Phantom

The CT ACR 464 phantom[65] is a cylindrical image quality phantom and is used for ACR CT accreditation testing. It contains four distinct modules (Figure 14) and can be used for testing various aspects of image quality:

- In-plane spatial resolution using the HU insert rods in module 1 (section 3.2.5.1)
- Z-direction spatial resolution using the interface between module 2 and module 3 (section 3.2.5.2)
- Noise magnitude and texture using the uniform section in module 3 (section 3.3)
- Quasi-linear task-based performance using the HU insert rods in module 1 and the uniform section in module 3 (section 3.4)
- Laser alignment accuracy using the fiducial markers in module 1 or 4 (section 6.1, Table 5a)
- Table indexing accuracy using the fiducial markers in module 1 or 4 (section 6.1, Table 5b)
- Image position accuracy using the fiducial markers in module 1 or 4 (section 6.1, Table 5c)
- Image thickness accuracy using the ramps in module 1 (section 6.1, Table 5d)

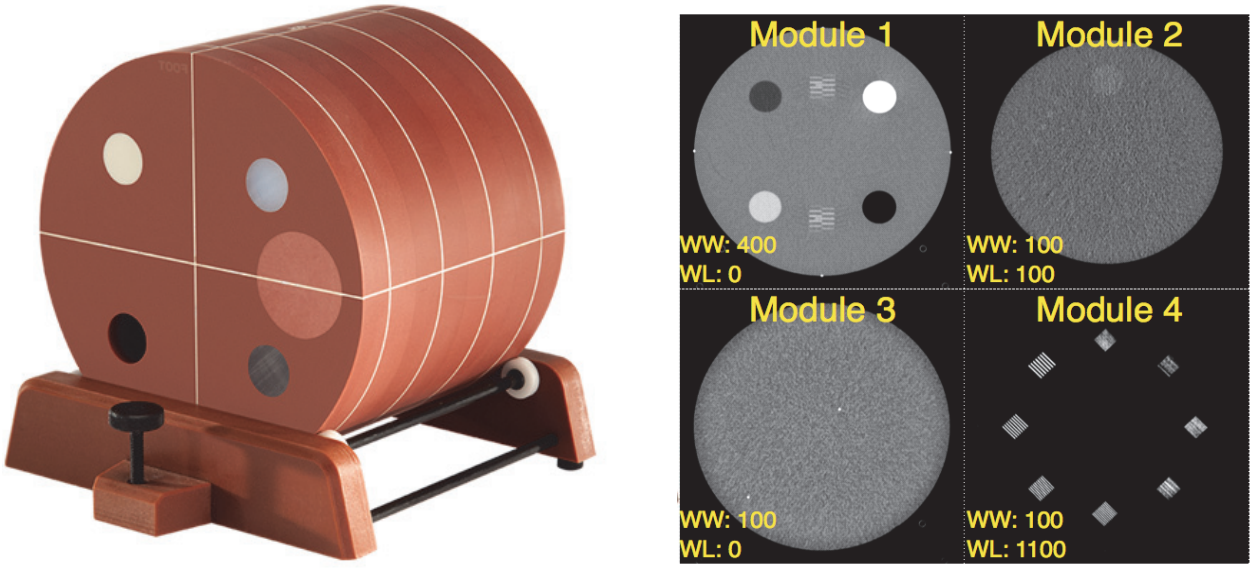

Figure 14. Photograph and representative CT images of the ACR CT phantom.

- Gantry tilt accuracy using one of several techniques (section 6.1, Table 5f)
- CT number accuracy using module 1 (section 6.3, Table 7a)
- CT number uniformity using module 3 (section 6.3, Table 7b)
- Artifact assessment using module 3 (section 6.3, Table 7c)
- Line-pair (high-contrast) resolution using module 4 (section 6.3, Table 7d)
- Low-contrast CNR using module 2 (section 6.3, Table 7e)
- Slice sensitivity profile using module 3 (section 6.3, Table 7g)

5.2.2 Mercury Phantom

The Mercury phantom is a cylindrical polyethylene phantom consisting of five sections of different diameters with tapered transitional sections (Figure 15).[22,23] Each sized section has two subsections: one uniform and the other containing cylindrical rods of varying materials (water, air, bone, polystyrene, and iodine). The phantom is designed to assess system noise, resolution, and detectability properties of the CT system as a function of patient size and detection task, and can be used for the following characterizations:

- TCM (both size adaptation and continuous adaptation) using the phantom sections of varying size (section 3.1)

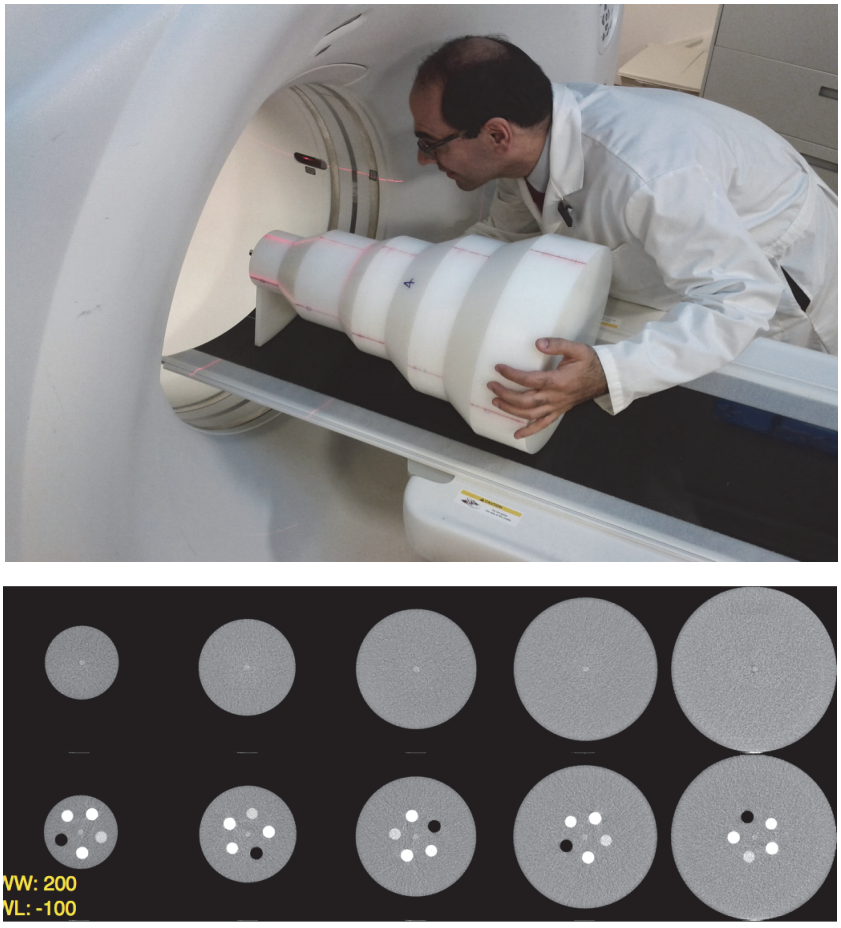

Figure 15. A rendering (top) of the Mercury phantom and example CT images of each phantom section (bottom). Each section is a different size and contains two subsections, a uniform section for noise characterization and a section with insert rods for characterizing resolution as a function of contrast. The noise and resolution measurements can then be combined to measure task-based performance as a function of phantom size.

- In-plane spatial resolution as a function of contrast, patient size, or image noise using the insert rods across variable size phantom sections (section 3.2.5.1)
- Z-direction spatial resolution using the slanted edge interface (section 3.2.5.2)
- Noise magnitude and texture as a function of patient size using the uniform phantom sections of varying size (section 3.3)
- Quasi-linear task-based performance as a function of patient size using the insert rods and the uniform sections of varying size (section 3.4)

5.2.3 Other Multi-size Phantoms

Besides the Mercury phantom, other multi-size phantom sets could be used for the following TCM tests:

- Size-adaptation of the TCM using at least two phantoms of different sizes (section 3.1.2.1)
- Continuous TCM adaptation using a phantom with continuous changes in water-equivalent diameter in the longitudinal direction (section 3.1.2.2)

Some examples of such phantoms are shown in Figure 16.

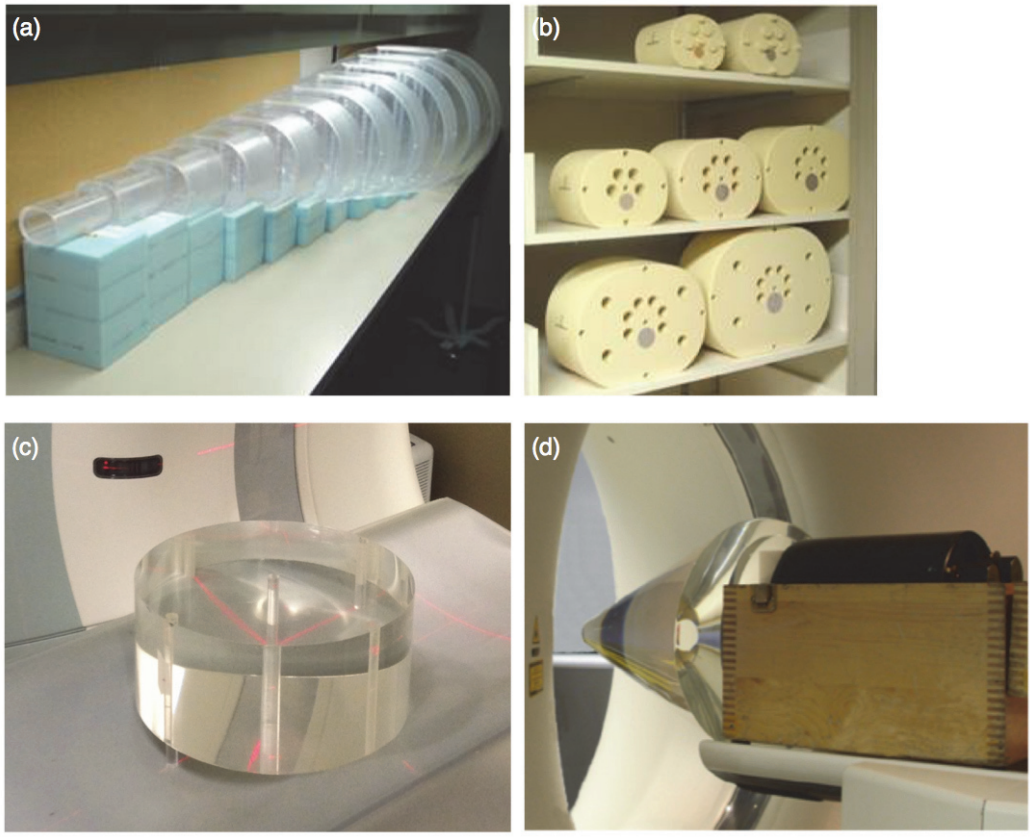

Figure 16. Photographs of various phantoms that can be used for TCM testing. Phantom sets of different sizes such as the collection of water phantoms (a), or CIRS abdominal phantoms (b) can be used for TCM size adaptation tests (see section 3.1.2.1). Phantoms with continuous longitudinal change in size. such as the CTDI phantom turned sideways (c), or the cone-shaped ImpACT phantom (d) can be used for TCM continuous adaptation tests (see section 3.1.2.1).

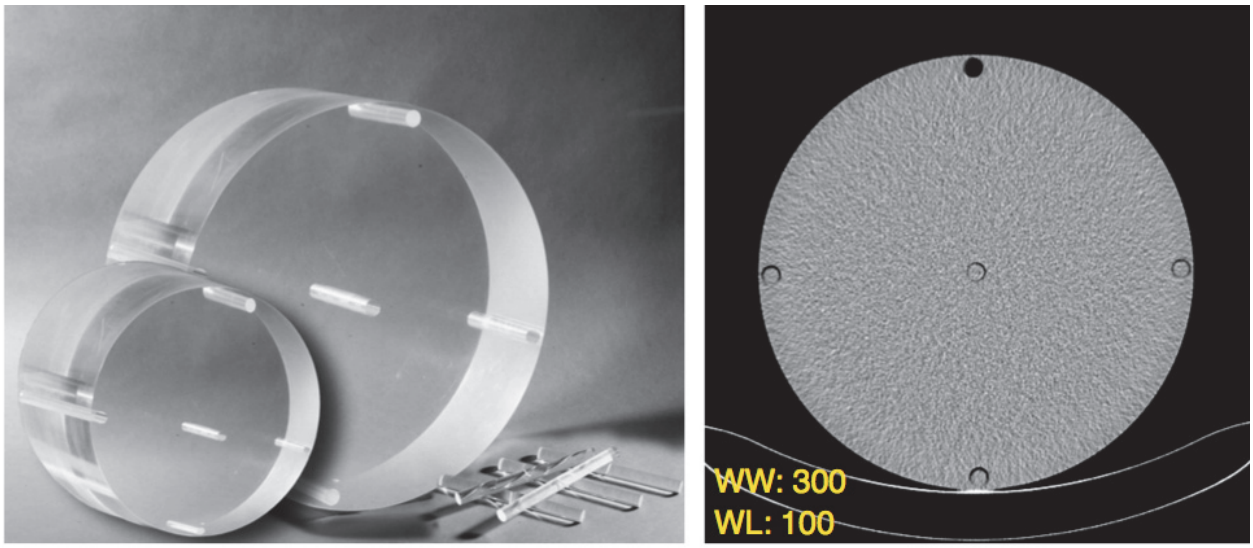

Figure 17. Photograph (left) and CT image (right) of the CTDI dosimetry phantoms.

5.2.4 CTDI Phantoms

CTDI phantoms are cylindrical acrylic phantoms used primarily for dosimetry purposes (Figure 17). This phantom family comes in two main diameters, 32 cm for body exams, and 16 cm for head and pediatric exams. Also a 10 cm "infant" phantom can be used for some tests. The phantoms have holes in the center and periphery for ion chamber measurements. The phantoms can be used for the following tests:

- Size-adaptation of the TCM using different-sized CTDI phantoms (section 3.1.2.1)
- Continuous TCM adaptation using a CTDI phantom placed sideways (section 3.1.2.2)
- Noise magnitude and texture (section 3.3)
- Displayed $CTDI_{vol}$ accuracy (section 6.2, Table 6h)
- Artifact assessment using images from any section of the phantom (section 6.3, Table 7c)

5.2.5 Low-contrast Detectability Phantoms

Spatial domain task-based performance requires phantoms with subtle low-contrast signals that can be imaged and presented to either computational or human observers (see section 3.5.4). Some examples of such phantoms are given in Figure 18 for the following test:

- Spatial domain task-based performance (section 3.5)

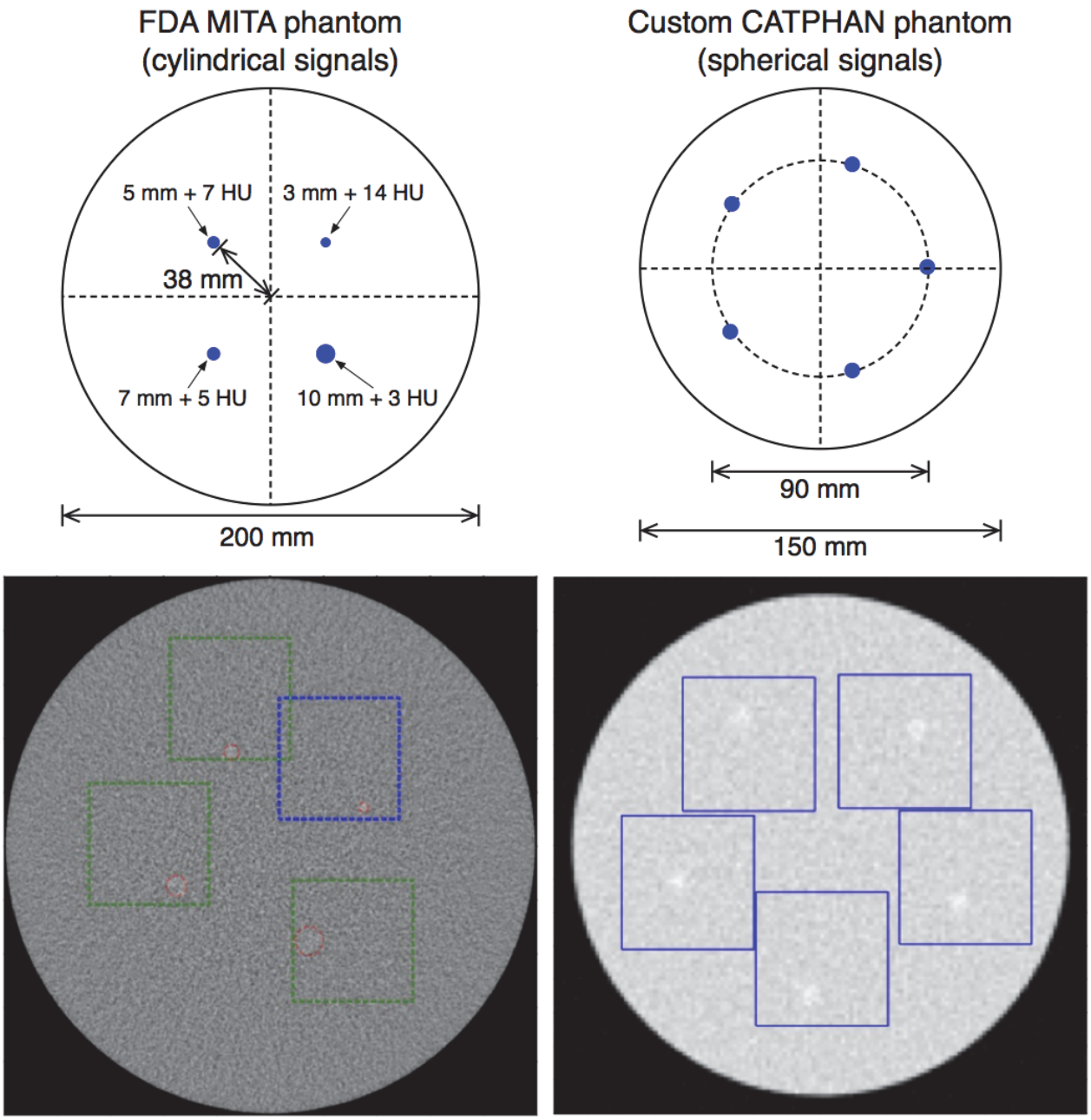

Figure 18. Design (top) and CT images (bottom) of the FDA MITA CT image quality phantom (left) and a custom Catphan image quality phantom (right). Both of these phantoms are designed to assess spatial domain task-based performance and include embedded low-contrast signals. The FDA MITA phantom has four 20-mm-long low-contrast rods of different diameter and contrast: 3 mm diameter, +14 HU contrast; 5 mm, +7 HU; 7 mm, +5 HU; and 10 mm, +3 HU. The custom Catphan phantom consists of signal modules with five identical spherical signals arranged symmetrically about the center. Modules with signals of different size or contrast could be fabricated and used. The spherical signals in this phantom allow for 3D image evaluations and volumetric signal searching as opposed to single slices.

5.2.6 Structured Phantoms

Phantoms with complex background structure (i.e., texture) can be used to test noise nonuniformity. Some examples of such phantoms are given in Figure 19 for the following tests:

- Noise nonuniformity index using repeated phantom images (section 3.3.5.2). Note that the noise nonuniformity index could also be calculated from non-structured (i.e., uniform) phantoms, but a structured phantom is preferred, especially if assessing iterative reconstruction.
- Noise inhomogeneity index using repeated images of a structured phantom (section 3.3.5.2).

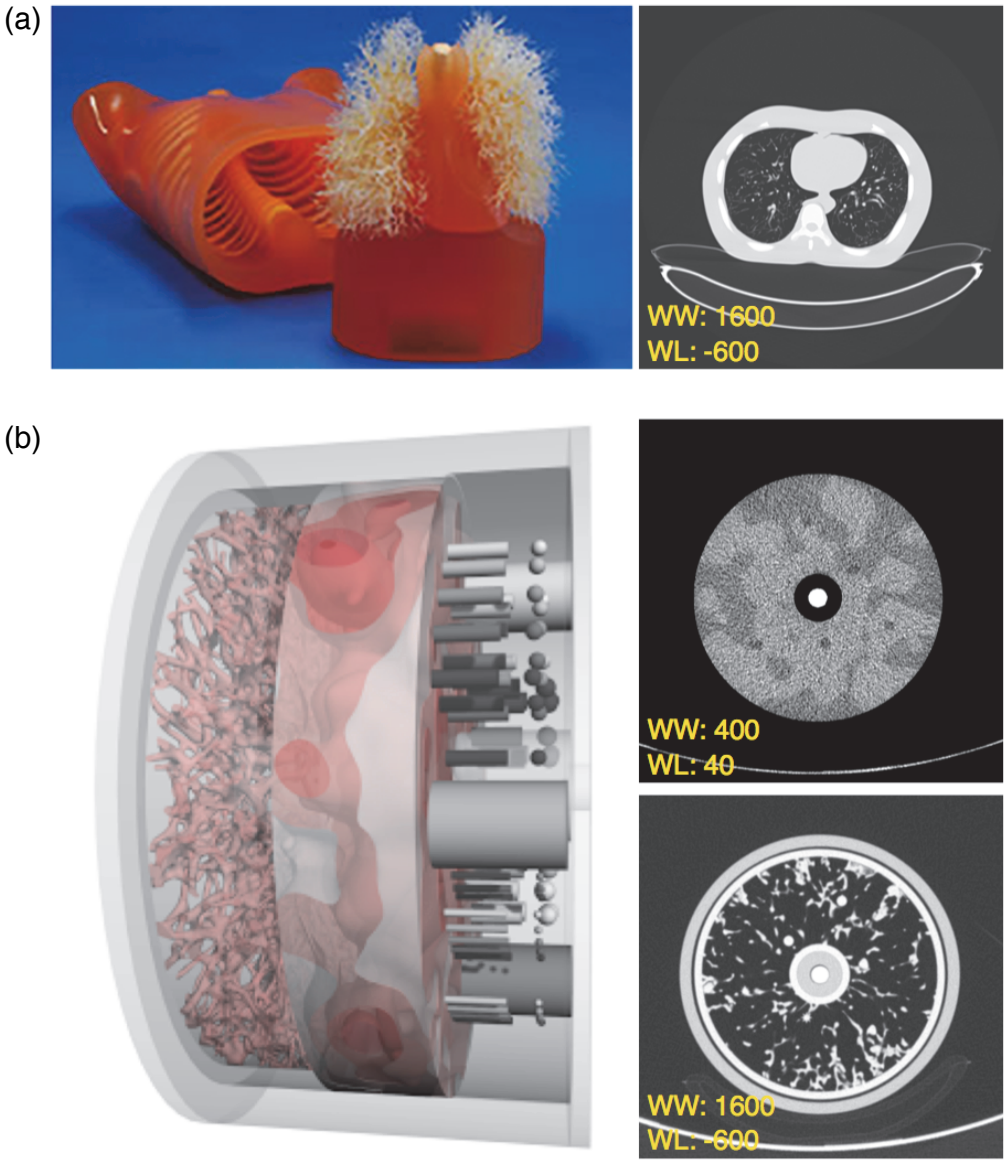

Figure 19. Examples of phantoms with complex anatomical structure suitable for assessment of noise nonuniformity. The Lungman chest phantom (a) is an anthropomorphic phantom containing detailed lung vasculature. The structured phantoms shown in (b) are 3D printed add-on modules meant to attach to the Mercury phantom platform. These extra modules contain simulated soft-tissue and lung texture.

5.2.7 Information for Commercially Available Phantoms

There are many phantoms on the market that may be used for CT operational performance evaluation. While some of these phantoms are not explicitly called out in this report, they may nonetheless be used to characterize many attributes of CT performance. We provide here a partial listing of commercially available phantoms that may be useful for performing some of the assessments described in this report. Inclusion of a phantom in this list does not imply endorsement by the AAPM, just as omission of a phantom in this list does not imply that it is deemed inappropriate for performing the assessments described in this report.

CIRS, Norfolk, Virginia

The CT Performance Phantom (Model 610) is an approximately 40-cm-long phantom that allows measurement of 10 CT performance parameters, including noise, detectability, alignment, slice thickness, uniformity, CT number linearity, and spatial resolution.
http://www.cirsinc.com/products/modality/31/aapm-ct-performance-phantom/

Low-Contrast Spherical Targets Phantom (Part 610-10) is a test object that contains low-contrast spheres. http://www.cirsinc.com/file/Products/610/610%20DS%20051315(1).pdf

Tissue Equivalent CT Dose Phantom (Model 007TE) is a family of four head phantoms, eight thorax phantoms, and eight abdominal phantoms representing different ages and sizes, some of which are shown in Figure 16b. These phantoms could be used to evaluate the performance of ATCM systems. http://www.cirsinc.com/products/modality/17/tissue-equivalent-ct-dose-phantoms/

Gammex, Inc., Middleton, Wisconsin (a Sun Nuclear Company)

The Mercury 4.0 Phantom is a 52-cm-long polyethylene phantom of varying diameters that allows for multiple assessments, including for ATCM systems, noise power spectrum, modulation and task transfer function, and detectability index.
https://www.sunnuclear.com/solutions/diagnostic/ct_solutions/mercury_phantom

The CT ACR Phantom (Model 464) is used for the American College of Radiology's CT Accreditation program and allows assessment of parameters such as positioning and alignment, CT number accuracy and uniformity, slice thickness, low-contrast detectability, and spatial resolution.
https://www.sunnuclear.com/solutions/diagnostic/ct_solutions/ct_acr_464_phantom

The Advanced iqModules are a four 20-cm-diameter, 4-cm-thick cylindrical phantoms that allow assessment of high-contrast resolution, low-contrast detectability, and uniformity over a wider range of specifications.
https://www.sunnuclear.com/solutions/diagnostic/ct_solutions/advanced-iqmodules

Kyoto Kagaku Co., LTD, Kyoto, Japan

The Multipurpose Chest Phantom "Lungman" is an anthropomorphic model of an adult male's chest that contains an anatomically complex representation of the pulmonary vessels. It could be used for ATCM performance evaluation or for noise texture analyses.
http://www.kyotokagaku.com/products/detail03/ph-1.html

Leeds Test Objects, North Yorkshire, United Kingdom

The CT Automatic Exposure Control phantom (CTAEC-50) includes six PMMA ellipses for assessment of ATCM performance. It is also available in an eleven ellipse configuration.
https://www.leedstestobjects.com/index.php/phantom/ct-aec-phantom/

The CT Image Quality phantom (CTIQ) is a PMMA cylinder with recesses in which to place various image quality test objects, including ones to test for CT number accuracy and uniformity,

noise, modulation transfer function, and spatial and contrast resolution.
https://www.leedstestobjects.com/index.php/phantom/to-ctiq-phantom/

QRM GmbH, Moehrendorf, Germany

QRM manufactures a wide range of image quality phantoms for the evaluation of parameters such as modulation transfer function and low- and medium-contrast resolution and spatial resolution. They also offer anthropomorphic CT phantoms of different sizes that may be suitable for ATCM evaluation.
http://www.qrm.de/content/products.htm

The Phantom Laboratory, Salem, New York

The name Catphan refers to a family of phantoms (Catphan 500, 600, 605, or 700) that offer performance characterization modules for the assessment of geometry, sensitometry, high-contrast resolution performance, low-contrast performance, and uniformity.
https://www.phantomlab.com/catphan-phantoms

The FDA MITA low-contrast detectability phantoms (Catphan 189 and 191), described in section 5.2.5, are designed for low-contrast imaging evaluations.
https://www.phantomlab.com/catphan-mita

The Automatic Tube Current Modulation (ATCM) phantom (CT228), is a 63-cm-long phantom composed of three ellipsoid sections having different long and short axis dimensions that was developed for the evaluation of ATCM systems.
https://www.phantomlab.com/atcm-phantom

5.3 Performance Evaluation Software

Several of the testing methods described in the Operational Performance section of this report (section 3) require involved calculations (e.g., quasi-linear task-based performance, section 3.4). Although it is possible (and acceptable) for any willing physicist to write and implement their own code to do these calculations, it is understood that many clinical physicists do not have the resources or time to do so. A number of free software packages are available to assist in these tasks. A few examples of such software packages are listed below.

- **imQuest**: CT image analysis tool used to extract tube current modulation profiles and measure spatial resolution, noise properties, and quasi-linear task-based performance based on the methods in this report (see sections 3.1–3.4). The tool is designed to work with the CT ACR 464 and Mercury phantoms, but could be used with any phantoms with similar features.
  http://deckard.mc.duke.edu/~samei/tg233.html
- **iMRMC**: Statistical analysis tool used to help do spatial domain task-based performance assessment (see section 3.5). The tool helps size and analyze multi-reader multi-case (MRMC) reader studies.
  https://github.com/DIDSR/iMRMC
- **iQModelo**: Tool including parametric statistical methods for ROC performance analysis of linear model observers (see section 3.5).
  https://github.com/DIDSR/IQmodelo

6. Appendix

This appendix provides further details in tabular form for the basic testing methods listed in Section 2. For complete descriptions of each test, readers are encouraged to seek the original source material.

6.1 Geometric Performance

Geometrical performance pertains to the basic aspects of the system functioning that are related to spatial reproducibility and accuracy. These are outlined in Table 5.

Tables 5a-f: Components of the assessment of geometrical performance

Table 5a. Laser Alignment Accuracy

Purpose	**To ensure that the laser alignment lights correctly indicate the scan position**		
	IEC Method	**ACR Method**	**IAEA Method**
Testing Devices	Thin absorber (e.g., a 1-mm-diameter wire)	Phantom that incorporates externally visible radiopaque fiducial markers or an image-center indication	Thin absorber (e.g., a 1-mm-diameter wire) placed on phantom
Setup	Test device is centered on internal laser	Align phantom in all three directions (x, y, z), and align all six degrees of freedom with the laser. Zero table location, ensure table motion is perpendicular to imaging plane.	Align on external laser, advance to internal laser, and verify the phantom is indeed aligned properly to internal laser.
Scan Protocol	Use narrowest beam and scan increment of <1 mm with a scan range of ±3 mm	Scan in axial mode using a reconstructed image width <2 mm or as thin as the scanner can produce. Use a technique appropriate to the phantom (e.g., adult abdomen)	Axial scan with thinnest image width (1 mm or less), ±3 mm scan range
Measurements	Select image location with the highest CT number	Determine location where markers are best visualized	Determine if full length of wire is visible within specifications
Qualitative Criteria	N/A	N/A	N/A
Quantitative Criteria	±2 mm	±2 mm	±5 mm acceptable (±1 mm achievable)
References	IEC standard 61223-3-5	ACR CT QC manual	IAEA Series No. 19
Notes	May also be tested using film with laser light position marked and automatic positioning of the table, if validated; repeat for external laser		May also be tested using film with laser light position marked and automatic positioning of the table, if validated; repeat for external laser

Table 5b. Table Indexing Accuracy

Purpose	**To ensure that the table moves as indicated**	
	IEC Method	**ACR Method**
Testing Devices	Ruler	A phantom with two sets of external fiducial markers of known separation
Setup	Place a table load ≤135 kg on table; attach ruler to fixed part of table; place mark on moving part of table and another mark adjacent to it on the ruler.	1) If possible, add weight to the tabletop to simulate the weight of an average patient. 2) Align phantom on first set of fiducial markers in the axial plane. 3) Zero the table position indicator.
Scan Protocol	1) Drive the table out a fixed distance. 2) Drive table back to original location setting. 3) Repeat for the opposite direction. 4) Repeat for CT condition of operation, in 10 mm increments up to 30 cm in both directions. 5) Repeat for stepped increments.	Move the table to the second set of external fiducial markers, record the table position, translate the table to full extension and return to the first set of fiducial markers, record the new table position.
Measurements	Using rulers from test setup: 1) Measure distance moved. 2) Measure distance between marks. 3) Repeat for opposite direction. 4) Repeat for CT condition of operation, in 10 mm increments up to 30 cm both directions. 5) Repeat for stepped increments.	See test setup/scan protocol
Qualitative Criteria	N/A	N/A
Quantitative Criteria	±1 mm	±1 mm
References	IEC standard 61223-3-5	ACR CT QC Manual
Notes	This test compares actual table motion to gantry-displayed distance. The table accuracy can be dependent on the patient weight and position relative to the contact point between the table and its driving mechanism (e.g., rollers). These factors are not varied in this test. This test does not determine if the image location corresponds to the actual table location. A mismatch here can result in gaps and overlaps, particularly in gated scans due to slippage in belts and rollers or errors in precise scanner measurement of table motion.	

Table 5c. Image Position Accuracy

Purpose	**To ensure that the prescribed image location indicated in a CT localizer radiograph correctly corresponds to the image position**		
	IEC Method	**ACR Method**	**IAEA Method**
Testing Devices	Thin absorber, e.g., 1-mm-diameter wire	A phantom that incorporates externally visible radiopaque fiducial markers or an image-center indication.	A suitable test tool can be easily fabricated from a PMMA slab at least 25 cm, and preferably >50 cm long. It can be shaped like a ruler and should have accurately placed markers at a set distance apart (e.g., 50 cm). Another alternative is to use a phantom or material block of a precisely known length.
Setup	Acquire a CT localizer radiograph of the test device.	Align phantom in all three directions, zero table location, and take a CT localizer radiograph. Magnify the image, if possible, and position a single image at the location of the radiopaque fiducial markers.	Place test tool along the long axis of the couch and acquire a CT localizer radiograph of the test device, making sure that the markers at each end of the tool are scanned.
Scan Protocol	Using the narrowest detector thickness and a table increment ≤1 mm, prescribe an axial scan directly covering the wire (test device) that covers a range of ±3 mm longitudinal to the test object	Perform an axial scan using a reconstructed scan width less than 2 mm or as thin as the scanner can produce in axial mode.	Locate the markers on the CT localizer radiograph and program two separate scans to create 1 mm (or the thinnest available) images directly over each of these markers.
Measurements	Select image location with the highest CT number	Determine location where markers are best visualized	The two acquired CT images of the markers, based on the CT localizer radiograph, should clearly display each marker.
Qualitative Criteria	N/A	N/A	N/A
Quantitative Criteria	±2 mm	±2 mm	±2 mm acceptable (±1 mm achievable)
References	IEC standard 61223-3-5	ACR CT QC manual	IAEA Series No. 19

Table 5d. Image Thickness Accuracy (Axial Mode)

Purpose	**To ensure that the nominal reconstructed image thickness is similar to the actual reconstructed image thickness**		
	IEC Method	**ACR Method**	**IAEA Method**
Testing Devices	One or (preferably) two ramps with known angles to the scan plane and with a linear attenuation coefficient of not less than that of aluminum and suitable for measuring all available tomographic section thicknesses.	A phantom with internal targets that allow the determination of reconstructed image thickness.	The test object typically contains one, or preferably two, thin metal inclined planes.
Setup	Align phantom inside gantry with the axis of rotation.	Align the phantom such that the phantom's image thickness determination targets are in the scan range.	Align the test device so that its axis coincides with the axis of rotation of the CT scanner and is centered in the field of view (FOV). Perform a CT localizer radiograph to confirm acceptable alignment and to define the tomographic plane.
Scan Protocol	Scan according to manufacturer instructions; for single-detector-row systems, all collimator settings should be measured; for multiple-detector-row systems, the maximum number of tomographic sections should be acquired for each collimator setting.	Using zero table increment, axial scan mode, and techniques adequate to allow unambiguous visualization of the targets (for most phantoms, 120 kV, 200 mAs is adequate), scan the phantom using each reconstructed image thickness used clinically.	Scan axially using clinical range of image widths. Ideally, all collimator settings should be tested and purchase specifications verified.
Measurements	For at least both the outer tomographic sections and one representative inner tomographic section, determine the maximum CT number (in HU) of each ramp, add to the background CT number, and divide by 2 to yield the full-width-at-half-maximum (FWHM). Set window level to FWHM and window width to 1 and measure length of visualized ramp. Average results if more than 1 ramp is used. Multiply by tangent of ramp angle.	If using the ACR CT accreditation program phantom, each line represents 0.5 mm thickness. Count each line that is at least 50% of the brightness of the brightest line. Divide the total number of lines by 2 to determine the reconstructed image thickness.	Determine FWHM (see IEC method).
Qualitative Criteria	N/A	N/A	N/A
Quantitative Criteria	±0.5 mm (for nominal image thicknesses ≤1 mm) ±50% of nominal thickness (for nominal image thicknesses 1–2 mm) ±1 mm (for nominal image thicknesses >2 mm)	±1.5 mm of nominal image thickness	±0.5 mm (for nominal image thicknesses ≤1 mm) ±50% of nominal thickness (for nominal image thicknesses 1–2 mm) ±1 mm (for nominal image thicknesses >2 mm)
References	IEC standard 61223-3-5	ACR CT QC manual	IAEA Series No. 19
Notes	A ramp with beads, disks, or wires may also be used.	This test was formerly required by the ACR accreditation program, but as of the 2017 CT QC manual, this test is no longer required or described.	When possible, scan protocols used for noise, dosimetry, and image width evaluation should be standardized and based on the clinical usage of the CT scanner.

Table 5e. Image Thickness Accuracy (Helical Mode)

Purpose	**To ensure that the nominal reconstructed image thickness is similar to the actual reconstructed image thickness**	
	IEC Method	**IAEA Method**
Testing Devices	Phantom containing a thin high contrast disc/bead (0.05 to 0.1 mm) with a linear attenuation coefficient of not less than that of aluminum.	The phantom should contain a thin (sub-millimeter thick) metal plate, or a sub-millimeter diameter air hole embedded in a uniform background cylinder (usually PMMA). The thickness of the metal plate or size of the air hole must be less than the smallest nominal image width that will be measured.
Setup	Align the test device such that the center of the test object aligns with the axis of rotation.	(1) Center the test object in the FOV on the table or secure it on a stand so that it is centered in the FOV; the metal foil insert needs to be parallel to the tomographic plane. (2) Perform a CT localizer radiograph and define the scan volume for helical scanning to ensure the metal foil is fully imaged.
Scan Protocol	Scan according to operation manual.	Determine helical image widths at a number of acquisition settings and reconstructed image thicknesses. Recommendations: (1) the reconstructed image width is the same width as the effective detector thickness; and (2) spot checks are made for larger reconstructed image widths. Investigate helical image widths under different pitch conditions.
Measurements	Reconstruct images at increments of 10% or less of the image thickness. Measure and record the mean CT number of the test object over the set of images and compute the FWHM	Place an ROI over the central portion of each reconstructed image corresponding to the position of the metal disc insert and measure and plot the average CT number for the range of images. Compute the FWHM using the CT number profile.
Qualitative Criteria	N/A	N/A
Quantitative Criteria	Vendor specifications	±0.5 mm (for nominal image thicknesses ≤1 mm) ±50% of nominal thickness (for nominal image thicknesses 1–2 mm) ±1 mm (for nominal image thicknesses >2 mm)
References	IEC standard 61223-3-5	IAEA Series No. 19

Table 5f. Gantry Tilt Accuracy

Purpose	**To ensure that the nominal gantry tilt is similar to the actual gantry tilt (IEC method) or that the gantry returns to a vertical position after being tilted (IAEA method)**	
	IEC Method	**IAEA Method**
Testing Devices	CR cassette	The following (or equivalent): a Lucite base and two Lucite pegs mounted on the base. The pegs are 5 cm high, 2.8 cm wide, and 25 cm apart. Vertical and horizontal holes 1 mm in diameter are drilled through the center of each peg, forming a cross inside the peg. Also a protractor or ruler is needed to measure angles.
Setup	On the CT table, sandwich the CR cassette between a pair of boxes so that it represents a sagittal plane with the laser beams running along its axes of symmetry.	Align the test device with the gantry lasers and verify that the device is aligned with the side vertical (axial) gantry lasers over the full range of vertical couch travel by moving the couch up and down. Tilt the gantry in both directions and then return to the vertical position.
Scan Protocol	In axial mode, expose the cassette three times with the gantry tilt set to zero, maximum superior, and maximum inferior positions using the thinnest available collimation setting.	N/A
Measurements	From the CR image, measure the angles of the dark lines relative to the vertical edge of the image.	The alignment of the laser QC device with vertical side (axial) gantry lasers should remain within 1 degree. Not explicitly stated in the IEC document: a protractor (or a ruler and trigonometry) can be used to measure this angle.
Qualitative Criteria	N/A	N/A
Quantitative Criteria	Accuracy of zero position ≤±1°	Within ±1 degree from vertical is acceptable and achievable.
References	IEC standard 61223-3-5	IAEA Series No. 19
Notes	In the case where an institution no longer employs CR reader technology, self-developing radiographic film can be substituted for a CR plate. An alternative technique is the following: While setting up for performing other measurements on any cylindrical phantom (e.g., ACR or CTDI), properly aligned on the table, acquire an image with the gantry at 0° and with the gantry tilted and measure the diameters anterior to posterior in the resultant images. Taking $\cos^{-1}$ of the diameter ratios characterizes the actual gantry tilt. It is also possible to use the ACR phantom's fiducial markers to test the gantry tilt angle by aligning the lasers to the center of module one, acquiring thin slice (<1 mm) images with the gantry tilted, recording the slice locations in which the top and bottom fiducial markers are visible, and using simple trigonometry to calculate a tilt angle.	

6.2 Radiation Output Performance

Radiation performance pertains to characterization of the radiation output of the CT system. These tests are outlined in Table 6.

Tables 6a-i: Components of the assessment of radiation performance of CT systems

Table 6a. Half-Value Layer (HVL)

Purpose	**To measure the half-value layer of the CT system's x-ray source and ensure that it is within regulatory limits**	
	Concentric Ring Method	**Stacked Sheet Method**
Testing Devices	CTDI ion chamber (pencil chamber with 10 cm active length) and electrometer; concentric 2 mm thick rings made of aluminum 1100.	CTDI ion chamber and electrometer; aluminum 1100 sheets.
Setup	CTDI ion chamber is placed at the gantry isocenter and centered along the z-direction. Axial CT scans are performed with the aluminum rings progressively nested to acquire CTDI data at increased filtration. The rings are placed on a low-attenuation stand such that the rings are centered about the chamber.	Park x-ray tube at a stationary position below isocenter. This could be achieved either using CT localizer radiograph scan, or with assistance of service engineer. The former requires a stationary support mechanism for the ion chamber other than patient table. CTDI ion chamber is placed at the gantry isocenter and centered along the z-direction. Aluminum sheets are gradually added on the CT gantry close to the x-ray tube.
Scan Protocol	Use relatively high mAs to ensure sufficient signal after several layers of aluminum are added. Perform measurements at each tube potential. Note that different bowtie filters could also be tested at each tube potential. Often the bowtie filter is determined by patient size selection for scan field of view.	
Measurements	Measure CTDI at each filtration level. HVL is calculated as the thickness of aluminum at which measured exposure is half of that without any aluminum filter (interpolation may be needed). Note that typical HVL range from approximately 4–8 mm aluminum and in some cases can be up to 10 mm aluminum.	
Qualitative Criteria	N/A	
Quantitative Criteria	CFR 21 specifies minimal HVL for general radiograph systems. HVL of CT system is usually higher than those specified in CFR 21. Recommend to use specifications from manufacturer.	
References	Kruger et al. 2000[66], IAEA Series No. 19, CFR 21	

Table 6b. Exposure Reproducibility

Purpose	**To ensure the radiation output of the system is consistent across repeated identical exposures**
Testing Devices	CTDI ion chamber (pencil chamber with 10 cm active length) and electrometer
Setup	CTDI ion chamber is placed at the gantry isocenter and centered along the z-direction. Test procedure is the same as CTDI measurement, but can be done in free air if preferred.
Scan Protocol	Use axial scan protocol for the two typical kV and mAs settings, representative of typical head and body techniques.
Measurements	CTDI measurements for the two typical kV and mAs settings, representative of typical head and body techniques. Repeat scans with each parameter setting.
Qualitative Criteria	N/A
Quantitative Criteria	The output should also be reproducible within a coefficient of variation of <0.10 (AAPM report 74) <20% of mean value of measurements taken.[10] For $CTDI_{free\text{-}in\text{-}air}$, each value shall be within ±10% of the mean of a set of 10 measurements (IEC).
References	AAPM report 74, EC report 162, IEC standard 61223-3-5

Table 6c. Exposure Time Reproducibility

Purpose	To ensure the exposure time is consistent across repeated identical exposures
Testing Devices	Dosimeters (ion chamber or other types) with time measurement capability.
Setup	Place the dosimeter at the gantry center.
Scan Protocol	Operate scanner in axial mode with no table translation.
Measurements	Take repeated measurements and record the exposure time for each nominal rotation time that is available on the scanner model.
Qualitative Criteria	N/A
Quantitative Criteria	Reproducibility of the exposure time should be within a coefficient of variation <0.05.
References	AAPM Report 74.

Table 6d. Exposure Linearity

Purpose	To ensure the radiation output of the system is linearly proportional to mAs
Testing Devices	CTDI ion chamber and electrometer
Setup	Put patient table just outside of scan range, place ion chamber on top of patient table, parallel to the gantry axis and centered both laterally and vertically (use stand or support if patient table is too low).
Scan Protocol	Operate scanner in axial mode with no table translation.
Measurements	For each tube potential, mAs can be varied in two ways (1) fix rotation time and change mA, (2) fix mA and change rotation time. Measure exposure at a range of mAs settings corresponding to typical clinical ranges.
Qualitative Criteria	N/A
Quantitative Criteria	Calculated CTDI/mAs for each parameter setting. Coefficient of linearity of CTDI/mAs between the mean of all values and any single value (absolute difference divided by sum) should be within 0.05.
References	AAPM Report 39.

Table 6e. Exposure Time Accuracy

Purpose	To ensure the nominal exposure time is similar to the actual exposure time
Testing Devices	Dosimeters (ion chamber or other types) with time measurement capability
Setup	Place the chamber at the gantry center.
Scan Protocol	Use axial protocol with no table translation
Measurements	Take measurements and record the exposure time for each nominal rotation time that is available on the scanner model.
Qualitative Criteria	N/A
Quantitative Criteria	For generators that display the selected time prior to the exposure, accuracy should be within ±5% (AAPM Report 74) Radiation termination shall occur within an interval that limits the total scan time to no more than 110 percent of its pre-set value (CFR 21 CT). Deviation from set value ≤±20%.
References	AAPM Report 74, CFR 21 CT, IEC standard 61223-3-5

Table 6f. Tube potential Accuracy

Purpose	To ensure the nominal tube potential is similar to the actual tube potential	
	Non-invasive Method	**Invasive Method**
Testing Devices	kV meter calibrated for CT	High voltage dividers.
Setup	Park the tube at top of the gantry, which can be achieved either in CT localizer radiograph mode or in service mode. Put table at lowest scan position, move tabletop into gantry opening, and place kV sensor on table. Align detector(s) to scan alignment light. If instrument detector is large or CT localizer radiograph mode is used instead of service mode, place the kV meter at bottom of gantry opening where the field size is greatest, with tabletop out of field.	Invasive test device (high voltage divider) directly measures the voltage of the generator. This is not recommended for routine test.
Scan Protocol	Set widest collimator setting and expose detector.	N/A
Measurements	Record kV values.	
Qualitative Criteria	N/A	
Quantitative Criteria	In absence of manufacturer's specifications, for both pulsed and non-pulsed generators, tube potential should be within ±2 kV of indicated values for all power levels.[9] ±5% nominal (Acceptable) and ±2% nominal (Achievable)[12].	
References	AAPM report 39 and AAPM report 74, IAEA Series No. 19	

Table 6g. Radiation Beam Profile

Purpose	To ensure the nominal radiation beam width is similar to the actual beam width
Testing Devices	External radiation detector. Area or long linear detectors such as CR plate, self-developing film, or optically stimulated luminescence (OSL) strip. Point detectors with temporal readout such as solid-state sensor or small ionization chamber with real time readout.
Setup	Preferred placement of radiation detector is in air at isocenter. However, this is not possible with all detectors: alternatively, a flat radiation attenuator shall be placed on the scanner table and then the radiation detector placed on top of the attenuator (e.g., CR plate placed on top of 15 cm acrylic sheet) and the table height adjusted to put radiation detector at isocenter.
Scan Protocol	Scan using each unique total beam collimation setting (N × T) available. • Area detectors (CR plate, film, long OSL) can be exposed with a single axial scan; • Point detectors (solid state sensor, small ionization chamber) require a helical scan through the entire beam width. • If necessary, available devices can transport small point detectors through the beam without having to move the table ("probe pullers"); consider their use for assessing scan modes/collimations with no helical scan option (e.g., a 320×0.5 mm beam width or other "volume scan" mode)
Measurements	From the radiation beam profile, calculate the full-width-at-half-maximum (FWHM) value.
Quantitative Criteria	Each manufacturer will have its own specifications for tolerances on FWHM of beam width at each collimation setting. Please consult manufacturer's documents. Generally narrower collimations require larger relative tolerances. ACR has suggested: FWHM should be accurate to within 3 mm or 30% of the total nominal collimated beam width (N × T), whichever is greater.
References	ACR CT QC manual
Notes	It is also possible to estimate the beam width using dosimetric measurements. For example, the CTDI pencil ion chamber can be used in conjunction with a radiopaque ring "mask" to estimate the dose per unit length for a given exposure setting (e.g., kV and mAs). A dose measurement at the collimation setting of interest is then divided by this dose per unit length to yield an estimate of the beam width.

Table 6h. Displayed $CTDI_{vol}$ Accuracy

Purpose	To ensure the displayed $CTDI_{vol}$ is similar to the actual $CTDI_{vol}$
Testing Devices	Calibrated electrometer and CTDI pencil ionization chamber (10 cm chamber); 16-cm (Head) CTDI dosimetry phantom, and 32-cm (Adult Body) CTDI dosimetry phantom.
Setup	1. Align the phantom (16 cm or 32 cm as appropriate for the scan protocol) such that the axis of the phantom is at the isocenter of the scanner and centered in all 3 planes. a. For head protocols, position the 16-cm phantom in the head holder or as heads are scanned clinically. b. For pediatric head protocols, place the 16-cm phantom directly on the scan table. c. For adult and pediatric abdomen protocols, place the 32-cm phantom directly on the scan table. 2. Connect the pencil chamber to the electrometer and insert the pencil chamber into the central hole in the phantom. Ensure that all other holes (those at 3, 6, 9, and 12 o'clock positions) are filled with acrylic rods.
Scan Protocol	The scans performed should include at least the following protocols: a. Adult Routine Brain b. Adult Routine Abdomen (70 kg) If Pediatric patients are scanned at the site, then these protocols should also be evaluated: c. Pediatric Routine Brain (1 year old) d. Pediatric Routine Abdomen (5 years old; 18–23 kg) For acceptance testing, the manufacturer generally provides the scan conditions for evaluating the $CTDI_{vol}$. For the ACR CT Accreditation Program, the site is required to use their clinical protocols; however, their clinical protocols may be different from those provided by the manufacturer. In addition, some accreditation or regulatory bodies may have specific requirements of the protocols used for evaluation of CT scanners or a facility (e.g., for comparison of accuracy). Using the appropriate protocol, acquire a single axial scan at the center of the phantom with no table increment. If the protocol is normally scanned helically, convert this to an axial scan while keeping the remaining technical parameters the same. All CTDI dose information must be acquired using axial scans.
Measurements and Calculations	For each protocol, 1. Record the $CTDI_{vol}$ reported by the scanner. 2. Position the phantom as described above. a. Place the dosimeter in the central position. b. Make one exposure in axial mode using the desired N × T configuration. If this configuration is not accessible in axial mode, use the N × T configuration most closely matching the desired value. c. Record the exposure value reported by electrometer (usually in units of mR in the U.S. and mGy elsewhere). d. Repeat the scan 2 more times and record the exposure. e. Repeat steps b-d above with the probe positioned at the 12 o'clock location (i.e., the periphery). 3. Averaging the 3 measurements done for each chamber position for each protocol, calculate the values (CTDI100,periphery) and (CTDI100,center) as CTDI100 = (f*CF*E*L)/(N*T), where f = conversion factor (8.7 mGy/R from exposure to dose in air, 1.0 mGy/mGy from air kerma to dose in air) CF = calibration or correction factor for electrometer E = average measured value (exposure or air kerma) L = active length of pencil ion chamber (typically 100 mm) N = actual number of data channels used during one axial acquisition T = width of each channel (N × T = nominal radiation beam width) 4. Calculate CTDIw = ((1/3) CTDI100,center) + ((2/3) CTDI100,periphery) 5. Calculate $CTDI_{vol}$ = $CTDI_w$ ((N x T) / I) = $CTDI_w$ / pitch, where I is the table increment per rotation (table speed). 6. Compare the measured $CTDI_{vol}$ to the values reported by the scanner.
Quantitative Criteria	Repeated measurements from year to year should not differ by more than 5%. Measured values should be within 20% of the values reported by the scanner; however, there may be conditions under which the manufacturer has specified wider tolerances (up to 50% in some cases) – these should be noted in the manufacturer's documentation. For ACR CT Accreditation Program, the measured $CTDI_{vol}$ values should not exceed the ACR CT Accreditation Program Reference Values, and must not exceed the ACR CT Accreditation Program Pass/Fail Values.

(continued)

Notes	**For acceptance testing**, the manufacturer generally provides the scan conditions for evaluating the $CTDI_{vol}$ and the tolerances for each test condition in the manufacturers' documentation. Check both the test conditions and the tolerances under those conditions. **For the ACR CT Accreditation Program**, the site is required to use their clinical protocols; however, these may be different from those provided by the manufacturer. If they are different from the manufacturer's test conditions, then the manufacturer's tolerances may not apply. It is **imperative to verify the phantom used by the manufacturer for the $CTDI_{vol}$ values reported by the scanner**. For all head scans (adult and pediatric) the 16-cm phantom is used. For adult body scans, the 32-cm dosimetry phantom is used. For pediatric body scans performed on some scanners, the 32-cm dosimetry phantom is used, while others use the 16-cm dosimetry phantom. Still, others base their choice of phantom on the patient size parameter selected. For example, "small" may use a 16-cm phantom, while "large" may use a 32-cm phantom. This variation is why it is important to verify which phantom is being used to report $CTDI_{vol}$ on the scanner for each protocol. While the international standard has been set so that the 16 cm phantom is used for all head (pediatric and adult) scans and the 32 cm phantom is used for all body (pediatric and adult) scans, the transition requires some scanners to be updated. Therefore, it is critical that the user always verify which phantom is being used when the scanner reports $CTDI_{vol}$. For **beam widths of greater than 40 mm,** the conventional $CTDI_{vol}$ calculations may substantially over-represent the dose to the phantom. Therefore, adjustments to the $CTDI_{vol}$ calculation may be necessary. These have been described in Geleijns et al.[67] If comparing to displayed values, then the method being used by the manufacturer should be described in detail, so consult the manufacturer's documentation. In multiple detector-row CT, $CTDI_{vol}$ is a function of detector configuration. Importantly, the detector configuration and total beam width used must match the collimation of the desired scanning protocol (N × T) as closely as possible. If the $N \times T$ value used for dosimetry does not exactly match the desired $CTDI_{vol}$ value, be sure to modify the table increment used in the calculation to yield the same pitch value as used in the scanning protocol.
References	ACR CT QC manual, IAEA Report 5, ICRU Report 87, Geleijns et al.[67]

Table 6i. CT Localizer Radiograph Dose

Purpose	To measure the exposure from the localizer radiograph
Testing Devices	Small ionization chamber, electrometer, foam block
Setup	CT localizer radiographs are essentially 2D radiographic images and are comparable to projection radiographs and not tomographic CT scans. Therefore, measure the radiation exposure from the CT localizer radiographs the same way exposure is measured for a single-plane projection radiography system. Place the ion chamber on top of a Styrofoam block on the CT table and position it so that the center of the chamber volume passes through the isocenter of the gantry during the acquisition of a projection radiograph.
Scan Protocol	Body CT localizer radiograph techniques (typically anteroposterior (AP) or PA). Acquire a projection view such that the ion chamber and its stem are in the approximate center of the image. To compare the measurements taken at the CT gantry isocenter to the entrance exposure measurements for a single plane radiograph, inverse-square corrections are applied to the collected exposure readings.
Measurements	Exposure in air values (mR) or air kerma (mGy)
Qualitative Criteria	No known criteria
Quantitative Criteria	No known criteria
References	O'Daniel et al. 2005[68] Schmidt et al. 2013[69]

6.3 Basic Image Quality Performance

Image quality performance pertains to the aspects of system performance that are related to characterization of reconstructed image. These are outlined in Table 7.

Tables 7a-g: Components of the assessment of imaging performance of CT systems

Table 7a. CT Number Accuracy

Purpose	**To ensure the CT numbers reported by the scanner are within an acceptable tolerance for known materials**		
	ACR Method	**IEC Method**	**IAEA Method**
Testing Devices	A phantom with inserts that provides >3 different CT numbers, including water and air values (e.g., ACR CT phantom module 1).		A water-filled test object (or phantom of a uniform material).
Setup	Align the phantom.		Center the phantom in the tomographic plane.
Scan Protocol	1. Use clinical protocols without AEC techniques such as TCM. 2. Use each kV available.		At acceptance, use manufacturer-specified acquisition parameters and phantoms. At commissioning, use a range of relevant kV, phantom sizes, reconstruction kernels, and scan modes (e.g., axial and helical).
Measurements	Record mean CT number for each target.		ROI diameter should be ~10% of phantom diameter. Record mean CT number.
Qualitative Criteria	N/A		N/A
Quantitative Criteria	water (–7 to 7 HU); air (–970 to –1005 HU); Teflon/bone (850 to 970 HU); polyethylene (–107 to –84 HU); acrylic (110 to 135 HU).	<±10 HU for water up to 30-cm diameter; different values will apply for other materials.	Water (acceptable: ±5 HU from baseline value; achievable: ±4 HU).
References	ACR CT QC manual	IEC standard 61223-3-5	IAEA Series No. 19
Notes	If using manufacturer's pass/fail criteria, scan technique must match manufacturer's recommendation. If using the ACR phantom, the tolerances above apply to 120 kV (or 130 kV if 120 is not available). For other kV settings, only the water and air tolerances above are applicable.		

Table 7b. CT Number Uniformity

Purpose	To ensure acceptable uniformity in CT numbers across the image field of view	
	ACR Method	**IAEA Method**
Testing Devices	Water phantom from manufacturer or ACR CT phantom module 3.	A water-filled test object (or phantom of a uniform material).
Setup	Align the phantom.	Center the phantom in the tomographic plane.
Scan Protocol	Use typical patient technique (kV, mA, rotation time, thickness, algorithm, and a relative smooth kernel), preferably matched to technologist's artifact analysis test.	At acceptance, use manufacturer-specified acquisition parameters and phantoms. At commissioning, use a range of relevant kV, phantom sizes, reconstruction kernels, and scan modes (e.g., axial and helical).
Measurements	ROI area should be ~1% of phantom area. Record mean CT number at center, 3, 6, 9, and 12 o'clock positions around phantom periphery.	ROI diameter should be ~10% of phantom diameter. Measure absolute difference of CT numbers between central and peripheral ROIs.
Qualitative Criteria	N/A	N/A
Quantitative Criteria	$CT_{peripheral}$-CT_{center} (acceptable: <7 HU; achievable: <5 HU).	$CT_{peripheral}$-CT_{center} (acceptable: <10 HU; achievable: <4 HU)
References	ACR CT QC manual	IAEA Series No. 19
Notes	Exclude bright peripheral ring by placing peripheral ROIs at a full ROI diameter from phantom edge.	

Table 7c. Artifact Assessment

Purpose	To ensure the images are free from artifacts	
	ACR Method	**IAEA Method**
Testing Devices	Water phantom from manufacturer or ACR CT phantom module 3.	A water-filled test object (or phantom of a uniform material).
Setup	Align the phantom.	Center the phantom in the tomographic plane.
Scan Protocol	Use typical patient technique (kV, mA, and rotation time) without automatic exposure control, preferably matched to technologist's artifact analysis test. Use thinnest axial images available, spanning the z-axis of the detector array.	At acceptance, use manufacturer-specified acquisition parameters and phantoms. At commissioning, use a range of relevant kV, phantom sizes, reconstruction kernels, and scan modes (e.g., axial and helical).
Measurements	Use appropriate width/level to optimize visibility of phantom material. Visually assess for artifacts (e.g., rings, streaks, lines, cupping, and capping). Be sure to inspect all acquired images.	Use appropriate width/level to optimize visibility of phantom material. Visually assess for artifacts (e.g., rings, streaks, lines, cupping, and capping). Be sure to inspect all acquired images.
Qualitative Criteria	Acceptable: no artifacts that have the potential to compromise diagnostic confidence. Achievable: no visible artifact.	Acceptable: no artifacts that have the potential to compromise diagnostic confidence. Achievable: no visible artifact.
Quantitative Criteria	N/A	N/A
References	ACR CT QC manual	IAEA Series No. 19
Notes	Best practice should also include a phantom with diameter >20 cm (e.g., large uniform phantoms from manufacturers or the 32-cm CTDI phantom).	

Table 7d. Line-pair (High-contrast) Resolution

Purpose	To estimate the limiting high-contrast (in-plane) spatial resolution of the system
Testing Devices	A phantom with high-contrast line-pair features (e.g., ACR, module 4)
Setup	Carefully align the phantom in all three directions at isocenter
Scan Protocol	Scan the phantom using the manufacturer's specified technique for comparison with manufacturer supplied specifications. Scan using clinical protocols to establish a baseline or to compare with previous baseline.
Measurements	Visually assess the line-pair images using an appropriate display window (e.g., WL= 1100 HU, WW= 100 HU for the ACR CT phantom). Record the highest spatial frequency for which the bars and spaces are distinctly visualized.
Qualitative Criteria	N/A
Quantitative Criteria	For ACR phantom, should be ≥6 lp/cm (7 lp/cm achievable) for standard soft-tissue reconstructions and ≥8 lp/cm (10 lp/cm achievable) for bone reconstructions.
References	ACR CT QC manual

Table 7e. Noise Magnitude

Purpose	To characterize the first-order noise properties of the CT system and to ensure the noise is consistent over time
Testing Devices	A phantom of a uniform material. For example, a water phantom or the ACR CT Accreditation Program module 3 can be used.
Setup	Center the phantom in the tomographic plane.
Scan Protocol	At acceptance, use manufacturer-specified acquisition parameters and phantoms. At commissioning, use a range of relevant kV, phantom sizes, reconstruction kernels, and scan modes (e.g., axial and helical).
Measurements	ROI diameter should be ~40% of phantom diameter. Record ROI standard deviation.
Qualitative Criteria	N/A
Quantitative Criteria	Acceptable: <±25% change from baseline value; Achievable: <±10%.
References	IAEA Series No. 19
Notes	Measurement of noise profile across image to investigate intra-image noise variation could also be useful at commissioning.

Table 7f. Low-Contrast Contrast-to-Noise Ratio (CNR)

Purpose	To estimate the low-contrast performance of the CT system and ensure that it is acceptable for diagnosis
Testing Devices	A phantom with low-contrast targets of known contrast (e.g., ACR CT Accreditation Program phantom module 2).
Setup	Align the phantom.
Scan Protocol	Use clinical protocols without automatic exposure control.
Measurements	Measure target contrast and background noise. Calculate contrast-to-noise ratio (CNR).
Qualitative Criteria	N/A
Quantitative Criteria	For ACR CT Accreditation Program phantom, visual performance for adult head and adult abdomen protocols must be ≥6-mm targets. CNR must be ≥1 for adult head and adult abdomen protocols, ≥0.7 for pediatric head protocols, and ≥0.5 for pediatric abdomen protocols.
References	ACR CT QC manual

Table 7g. Slice Sensitivity Profile (SSP)

Purpose	To estimate the high-contrast z-direction spatial resolution of the system
Testing Devices	A phantom containing an embedded high-contrast small spherical feature (e.g., ACR CT Accreditation Program phantom module 3).
Setup	Align the phantom.
Scan Protocol	Use clinical protocols in which z-direction resolution is thought to be important (e.g., temporal bone). Images should be reconstructed with the slice thickness used clinically for the protocol of interest, but the spacing between slices should be minimized in order to properly sample the SSP. The scan range should be centered about the phantom's high-contrast feature and be long enough to include the tales of the SSP.
Measurements	In all image slices surrounding the spherical feature, place an ROI around the feature and record the maximum CT number. Plot the maximum CT number as a function of slice position and record the FWHM of the SSP.
Qualitative Criteria	N/A
Quantitative Criteria	N/A
References	Greene and Rong 2014[26]

7. References

1. Zhu, T.C., A. Ahnesjo, K.L. Lam, X.A. Li, C.M. Ma, J.R. Palta, M.B. Sharpe, B. Thomadsen, and R.C. Tailor, "Report of AAPM Therapy Physics Committee Task Group 74: in-air output ratio, Sc, for megavoltage photon beams." *Med Phys*, 2009. **36**(11): p. 5261–91.
2. Huda, W. and F.A. Mettler, "Volume CT dose index and dose-length product displayed during CT: what good are they?" *Radiology*, 2011. **258**(1): p. 236–42.
3. McCollough, C.H., S. Leng, L. Yu, D.D. Cody, J.M. Boone, and M.F. McNitt-Gray, "CT dose index and patient dose: they are not the same thing." *Radiology*, 2011. **259**(2): p. 311–6. PMCID:PMC3079120.
4. The Association of Electrical Equipment and Medical Imaging Manufacturers, *MITA Smart Dose CT XR-29: Standard Attributes on CT Equipment Related to Dose Optimization and Management*, 2013.
5. The Association of Electrical Equipment and Medical Imaging Manufacturers, *NEMA XR 25-2010: Computed Tomography Dose Check*, 2010.
6. National Committee on Radiation Protection, *NCRP Report No. 147: Structural Shielding Design for Medical X-Ray Imaging Facilities*, 2004.
7. American College of Radiology. *CT Accredidation Program*. 2014 [cited 2014 July 16]; Available from: http://www.acr.org/Quality-Safety/accreditation/CT.
8. Zhu, T.C., A. Ahnesjo, K.L. Lam, X.A. Li, C.M. Ma, J.R. Palta, M.B. Sharpe, B. Thomadsen, R.C. Tailor, and A.T.P.C.T. Group, "Report of AAPM Therapy Physics Committee Task Group 74: in-air output ratio, Sc, for megavoltage photon beams." *Med Phys*, 2009. **36**(11): p. 5261–91.
9. American Association of Physicists in Medicine, *AAPM Report No. 39: Specification and acceptance testing of computed tomography scanners*, 1993.
10. European Commission, *Criteria for Acceptability of Medical Radiological Equipment used in Diagnostic Radiology, Nuclear Medicine and Radiotherapy*, 2012.
11. Food and Drug Administration, Code Of Federal Regulations Part 120: Performance Standards for Ionizing Radiation Emitting Products, 2013.
12. International Atomic Energy Agency, *Quality Assurance Programme for Computed Tomography: Diagnostic and Therapy Applications*, 2012.
13. International Atomic Energy Agency, *Status of Computed Tomography Dosimetry for Wide Cone Beam Scanners*, 2011.
14. International Commission on Radiation Units and Measurements, *ICRU Report No. 87: Radiation dose and image-quality assessment in computed tomography*, 2012.
15. International Electrotechnical Commission, *Evaluation and routine testing in medical imaging departments – Part 3–5: Acceptance tests – Imaging performance of computed tomography X-ray equipment*, 2004.
16. Gang, G.J., J. Lee, J.W. Stayman, D.J. Tward, W. Zbijewski, J.L. Prince, and J.H. Siewerdsen, "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance." *Med Phys*, 2011. **38**(4): p. 1754–68. PMCID: PMC3069989
17. Gifford, H.C., M.A. King, P.H. Pretorius, and R.G. Wells, "A comparison of human and model observers in multislice LROC studies." *IEEE Trans Med Imaging*, 2005. **24**(2): p. 160–9.
18. Leng, S., L. Yu, Y. Zhang, R. Carter, A.Y. Toledano, and C.H. McCollough, "Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain." *Med Phys*, 2013. **40**(8): p. 081908. PMCID:PMC3724792.
19. Christianson, O., J.J. Chen, Z. Yang, G. Saiprasad, A. Dima, J.J. Filliben, A. Peskin, C. Trimble, E.L. Siegel, and E. Samei, "An Improved Index of Image Quality for Task-based Performance of

CT Iterative Reconstruction across Three Commercial Implementations." *Radiology*, 2015. **275**(3): p. 725–34.

20. McCollough, C., D.M. Bakalyar, M. Bostani, S. Brady, K. Boedeker, J.M. Boone, H.H. Chen-Mayer, O.I. Christianson, S. Leng, B. Li, M.F. McNitt-Gray, R.A. Nilsen, M.P. Supanich, and J. Wang, "Use of Water Equivalent Diameter for Calculating Patient Size and Size-Specific Dose Estimates (SSDE) in CT," AAPM Report 220, 2014.
21. Medicines and Healthcare products Regulatory Agency, *MHRA Report 05016: CT scanner automatic exposure control systems*, 2005.
22. Solomon, J.B., J. Wilson, and E. Samei, "Characteristic image quality of a third generation dual-source MDCT scanner: noise, resolution, and detectability." *Med Phys*, 2015. **42**(8): p. 4941–53.
23. Wilson, J.M., O.I. Christianson, S. Richard, and E. Samei, "A methodology for image quality evaluation of advanced CT systems." *Med Phys*, 2013. **40**(3): p. 031908. PMID: 23464323
24. Chen, B., O. Christianson, J.M. Wilson, and E. Samei, "Assessment of volumetric noise and resolution performance for linear and nonlinear CT reconstruction methods." *Med Phys*, 2014. **41**(7): p. 071909.
25. Richard, S., D.B. Husarik, G. Yadava, S.N. Murphy, and E. Samei, "Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms." *Med Phys*, 2012. **39**(7): p. 4115–22. PMID: 22830744.
26. Greene, T.C. and X.J. Rong, "Evaluation of techniques for slice sensitivity profile measurement and analysis." *J Appl Clin Med Phys*, 2014. **15**(2): p. 4042.
27. Maidment, A.D. and M. Albert, "Conditioning data for calculation of the modulation transfer function." *Med Phys*, 2003. **30**(2): p. 248–53.
28. Cruz-Bastida, J.P., D. Gomez-Cardona, K. Li, H. Sun, J. Hsieh, T.P. Szczykutowicz, and G.H. Chen, "Hi-Res scan mode in clinical MDCT systems: Experimental assessment of spatial resolution performance." *Med Phys*, 2016. **43**(5): p. 2399. PMCID:PMC4841803.
29. Yu, L., T.J. Vrieze, S. Leng, J.G. Fletcher, and C.H. McCollough, "Technical Note: Measuring contrast- and noise-dependent spatial resolution of an iterative reconstruction method in CT using ensemble averaging." *Med Phys*, 2015. **42**(5): p. 2261–7. PMCID:PMC4401802
30. Li, K., J. Tang, and G.H. Chen, "Statistical model based iterative reconstruction (MBIR) in clinical CT systems: experimental assessment of noise performance." *Med Phys*, 2014. **41**(4): p. 041906. PMCID:PMC3978426.
31. Solomon, J. and E. Samei, "Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE." *Med Phys*, 2014. **41**(9): p. 091908.
32. Boedeker, K.L., V.N. Cooper, and M.F. McNitt-Gray, "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**(14): p. 4027–46.
33. Solomon, J.B., O. Christianson, and E. Samei, "Quantitative comparison of noise texture across CT scanners from different manufacturers." *Med Phys*, 2012. **39**(10): p. 6048–55.
34. Solomon, J., A. Ba, F. Bochud, and E. Samei, "Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms." *Med Phys*, 2016. **43**(12): p. 6497.
35. Li, K., J. Garrett, Y. Ge, and G.H. Chen, "Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance." *Med Phys*, 2014. **41**(7): p. 071911. PMCID:PMC4106476
36. Solomon, J., J. Wilson, and E. Samei, "Characteristic image quality of a third generation dual-source MDCT scanner: Noise, resolution, and detectability." *Med Phys*, 2015. **42**(8): p. 4941–53.

37. Siewerdsen, J.H., I.A. Cunningham, and D.A. Jaffray, "A framework for noise-power spectrum analysis of multidimensional images." *Med Phys*, 2002. **29**(11): p. 2655–71.
38. International Commission on Radiation Units and Measurements, *ICRU Report 54: Medical Imaging–The Assessment of Image Quality*, 1995.
39. Solomon, J., D. Marin, K. Roy Choudhury, B. Patel, and E. Samei, "Effect of Radiation Dose Reduction and Reconstruction Algorithm on Image Noise, Contrast, Resolution, and Detectability of Subtle Hypoattenuating Liver Lesions at Multidetector CT: Filtered Back Projection versus a Commercial Model-based Iterative Reconstruction Algorithm." *Radiology*, 2017: p. 161736.
40. Solomon, J., A. Mileto, J.C. Ramirez-Giraldo, and E. Samei, "Diagnostic Performance of an Advanced Modeled Iterative Reconstruction Algorithm for Low-Contrast Detectability with a Third-Generation Dual-Source Multidetector CT Scanner: Potential for Radiation Dose Reduction in a Multireader Study." *Radiology*, 2015. **275**(3): p. 735–45.
41. Samei, E., M.J. Flynn, and W.R. Eyler, "Simulation of subtle lung nodules in projection chest radiography." *Radiology*, 1997. **202**(1): p. 117–24.
42. Chen, B. and E. Samei. "Development of a phantom-based methodology for the assessment of quantification performance in CT." In *SPIE Medical Imaging*, 2013. Orlando, FL. **8668** 86681E-86681E-7.
43. Popescu, L.M. and K.J. Myers, "CT image assessment by low contrast signal detectability evaluation with unknown signal location." *Med Phys*, 2013. **40**(11): p. 111908.
44. Zhang, Y., S. Leng, L. Yu, R.E. Carter, and C.H. McCollough, "Correlation between human and model observer performance for discrimination task in CT." *Phys Med Biol*, 2014. **59**(13): p. 3389–404. PMCID:PMC4057982.
45. Solomon, J.B. and E. Samei, "Correlation between human detection accuracy and observer model-based image quality metrics in computed tomography." *J Med Imaging*, 2016. **3**: p. 12.
46. Abbey, C.K. and H.H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability." *J Opt Soc Am A Opt Image Sci Vis*, 2001. **18**(3): p. 473–88. PMCID:PMC2943344.
47. Yu, L., S. Leng, L. Chen, J.M. Kofler, R.E. Carter, and C.H. McCollough, "Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms." *Med Phys*, 2013. **40**: p. 041908.
48. Barrett, H.H., K.J. Myers, C. Hoeschen, M.A. Kupinski, and M.P. Little, "Task-based measures of image quality and their relation to radiation dose and patient risk." *Phys Med Biol*, 2015. **60**(2): p. R1–R75. PMCID:PMC4318357.
49. Favazza, C.P., A. Ferrero, L. Yu, S. Leng, K.L. McMillan, and C.H. McCollough, "Use of a channelized Hotelling observer to assess CT image quality and optimize dose reduction for iteratively reconstructed images." *J Med Imaging* (Bellingham), 2017. **4**(3): p. 031213. PMCID: PMC5624940.
50. Popescu, L.M., "Nonparametric ROC and LROC analysis." *Med Phys*, 2007. **34**(5): p. 1556–64.
51. Wunderlich, A. and F. Noo, "A nonparametric procedure for comparing the areas under correlated LROC curves." *IEEE Trans Med Imaging*, 2012. **31**(11): p. 2050-61. PMCID: PMC3619029.
52. Gallas, B.D., "One-shot estimate of MRMC variance: AUC." *Acad Radiol*, 2006. **13**(3): p. 353–62.
53. Popescu, L.M., "Nonparametric signal detectability evaluation using an exponential transformation of the FROC curve." *Med Phys*, 2011. **38**(10): p. 5690–702.

54. Vaishnav, J.Y., W.C. Jung, L.M. Popescu, R. Zeng, and K.J. Myers, "Objective assessment of image quality and dose reduction in CT iterative reconstruction." *Med Phys*, 2014. **41**(7): p. 071904.
55. Becchetti, M.F., J.B. Solomon, W.P. Segars, and E. Samei. "Synthesized interstitial lung texture for use in anthropomorphic computational phantoms." *Proc SPIE*, 2016. **9783:** 97835Z-97835Z-5.
56. Bolch, W., C. Lee, M. Wayson, and P. Johnson, "Hybrid computational phantoms for medical dose reconstruction." *Radiat Environ Biophys*, 2010. **49**(2): p. 155–68. PMCID:PMC2855752.
57. Lee, C., D. Lodwick, J. Hurtado, D. Pafundi, J.L. Williams, and W.E. Bolch, "The UF family of reference hybrid phantoms for computational radiation dosimetry." *Phys Med Biol*, 2010. **55**(2): p. 339–63. PMCID:PMC2800036.
58. Norris, H., Y. Zhang, J. Bond, G.M. Sturgeon, A. Minhas, D.J. Tward, J.T. Ratnanather, M.I. Miller, D. Frush, E. Samei, and W.P. Segars, "A set of 4D pediatric XCAT reference phantoms for multimodality research." *Med Phys*, 2014. **41**(3): p. 033701. PMCID:PMC3987726.
59. Segars, W.P., H. Norris, G.M. Sturgeon, Y. Zhang, J. Bond, A. Minhas, D.J. Tward, J.T. Ratnanather, M.I. Miller, D. Frush, and E. Samei, "The development of a population of 4D pediatric XCAT phantoms for imaging research and optimization." *Med Phys*, 2015. **42**(8): p. 4719–26. PMCID:PMC4506297.
60. Solomon, J., A. Ba, A. Diao, J. Lo, E. Bier, F. Bochud, M. Gehm, and E. Samei. "Design, fabrication, and implementation of voxel-based 3D printed textured phantoms for task-based image quality assessment in CT." *SPIE Medical Imaging*, 2016. **9783** 978328-978328-11.
61. Ma, C., L. Yu, B. Chen, C. Favazza, S. Leng, and C. McCollough, "Impact of number of repeated scans on model observer performance for a low-contrast detection task in computed tomography." *J Med Imaging* (Bellingham), 2016. **3**(2): p. 023504. PMCID:PMC4886187.
62. Tseng, H.W., J. Fan, M.A. Kupinski, P. Sainath, and J. Hsieh, "Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers." *Med Phys*, 2014. **41**(7): p. 071910.
63. Winslow, J., Y. Zhang, and E. Samei, "A method for characterizing and matching CT image quality across CT scanners from different manufacturers." *Med Phys*, 2017. **44**(11): p. 5705–5717.
64. Zhang, Y., C. Smitherman, and E. Samei, "Size-specific optimization of CT protocols based on minimum detectability." *Med Phys*, 2017. **44**(4): p. 1301–1311.
65. McCollough, C.H., M.R. Bruesewitz, M.F. McNitt-Gray, K. Bush, T. Ruckdeschel, J.T. Payne, J.A. Brink, and R.K. Zeman, "The phantom portion of the American College of Radiology (ACR) computed tomography (CT) accreditation program: practical tips, artifact examples, and pitfalls to avoid." *Med Phys*, 2004. **31**(9): p. 2423–42.
66. Kruger, R.L., C.H. McCollough, and F.E. Zink, "Measurement of half-value layer in x-ray CT: a comparison of two noninvasive techniques." *Med Phys*, 2000. **27**(8): p. 1915–9.
67. Geleijns, J., M. Salvado Artells, P.W. de Bruin, R. Matter, Y. Muramatsu, and M.F. McNitt-Gray, "Computed tomography dose assessment for a 160 mm wide, 320 detector row, cone beam CT scanner." *Phys Med Biol*, 2009. **54**(10): p. 3141–59. PMCID:PMC2948862.
68. O'Daniel, J.C., D.M. Stevens, and D.D. Cody, "Reducing radiation exposure from survey CT scans." *AJR Am J Roentgenol*, 2005. **185**(2): p. 509–15.
69. Schmidt, B., N. Saltybaeva, D. Kolditz, and W.A. Kalender, "Assessment of patient dose from CT localizer radiographs." *Med Phys*, 2013. **40**(8): p. 084301.