

Variability in image quality and radiation dose within and across 97 medical facilities

Taylor B. Smith^{a,b,c,*}, Shuaiqi Zhang,^d Alaattin Erkanli,^d
Donald Frush,^e and Ehsan Samei^{a,b,c}

^aDuke University, Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Durham, North Carolina, United States

^bDuke University, Medical Physics Graduate Program, Durham, North Carolina, United States

^cDuke University Medical Center, Durham, North Carolina, United States

^dDuke University School of Medicine, BERD Methods Core, Department of Biostatistics and Bioinformatics, Durham, North Carolina, United States

^eStanford University, Lucile Salter Packard Children's Hospital, Stanford, California, United States

Abstract

Purpose: To characterize variability in image quality and radiation dose across a large cohort of computed tomography (CT) examinations and identify the scan factors with the highest influence on the observed variabilities.

Approach: This retrospective institutional-review-board-exempt investigation was performed on 87,629 chest and abdomen-pelvis CT scans acquired for 97 facilities from 2018 to 2019. Images were assessed in terms of noise, resolution, and dose metrics (global noise, frequency in which modulation transfer function is at 0.50, and volumetric CT dose index, respectively). The results were fit to linear mixed-effects models to quantify the variabilities as affected by scan parameters and settings and patient characteristics. A list of factors, ranked by t -value with $p < 0.05$, was ascertained for each of the six mixed effects models. A type III p -value test was used to assess the influence of facility.

Results: Across different facilities, image quality and dose were significantly different ($p < 0.05$), with little correlation between their mean magnitudes and consistency (Pearson's correlation coefficient < 0.34). Scanner model, slice thickness, recon field-of-view and kernel, mAs, kVp, patient size, and centering were the most influential factors. The two body regions exhibited similar rankings of these factors for noise (Spearman's correlation coefficient = 0.76) and dose (Spearman's correlation coefficient = 0.86) but not for resolution (Spearman's correlation coefficient = 0.52).

Conclusions: Clinical CT scans can vary in image quality and dose with broad implications for diagnostic utility and radiation burden. Average scan quality was not correlated with interpatient scan-quality consistency. For a given facility, this variability can be quite large, with magnitude differences across facilities. The knowledge of the most influential factors per body region may be used to better manage these variabilities within and across facilities.

© 2021 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.8.5.052105](https://doi.org/10.1117/1.JMI.8.5.052105)]

Keywords: Image quality; radiation dose; patient-specific; computed tomography.

Paper 20257SSRR received Sep. 25, 2020; accepted for publication Apr. 13, 2021; published online May 8, 2021.

*Address all correspondence to Taylor B. Smith, taylor.smith@duke.edu

1 Introduction

Computed tomography (CT) has established itself in the clinic as an indispensable tool for medical diagnosis across all ages. Even so, the practice of CT imaging is an inherently heterogeneous operation as every CT scan may have different quality (e.g., Fig. 1), which can translate to differences in diagnostic utility. This variability is due to variation in the way different imaging facilities practice CT imaging.¹⁻⁵ For a given facility, there are often variations in makes and models of the equipment. Further, these are differences in the imaging protocols used. Ideally, conscious decisions and philosophies regarding sufficient image quality or dose govern the practice. However, many choices are made on an *ad hoc* basis. The diversity of patient anatomy and physiology only adds to this landscape of heterogeneity, creating an intrafacility variability in image quality and radiation dose at a given facility across its patient population.^{1,6} Further without a systematic standard, there exist systematic differences from facility to facility, influencing image quality and dose associated with clinical images across facilities. This interfacility variability of quality is one aspect that leads to overall variability in CT imaging as practiced today. Variability in the quality of images can influence the efficacy and confidence of detection and diagnostic tasks.⁷ Thus there is a need to understand and manage this variability to improve consistency of care and overall utility of CT imaging.

The first step toward improving consistency in diagnostic value is understanding the scope and magnitude of the variability. Initiatives like the American College of Radiology's "Dose Index Registry" emphasize the importance of quantifying the variability of current practice in terms of CT radiation dose and its derivatives.³ The multitude of dose monitoring solutions further shows the widespread recognition of this need in clinical communities. Although radiation burden incurred by CT is important, it is not the entire story. Image quality is perhaps as

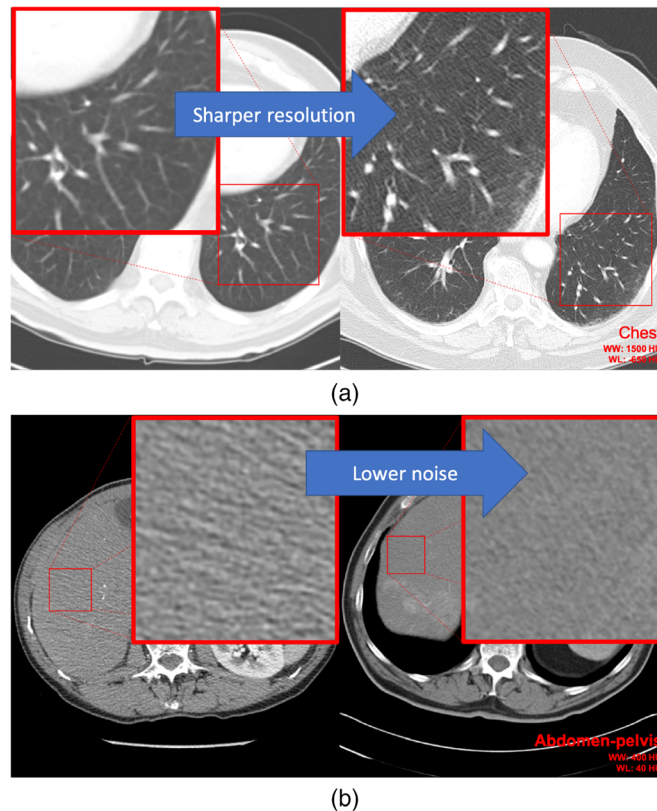


Fig. 1 Four slices from different CT scans show how noise and resolution differ between different scans—with differences in (a) image resolution and (b) image noise showcased. In the chest exams (a), the sharpness difference is evident in the depiction of vessels. In the abdomen-pelvis exams (b), the difference is evident in the noisiness of the patients' livers.

much or more relevant than dose and more clearly attributable to enhancing care. Aspects of image quality such as noise and resolution should likewise be managed to improve consistency. Recent studies have shown how noise and resolution can be efficiently and effectively assessed in large numbers of individual patient exams.^{8,9} These studies have been important first steps toward understanding the variability of patient imaging, but they have been limited either in (1) their scope and number of facilities included or (2) their focus on only dose or noise, rather than dose, noise, and resolution simultaneously across a large population.^{1,6,10–12}

This investigation aimed to characterize the current state of variability in both dose and image quality in clinical CT imaging at multiple facilities by assessing each on 87,629 CT scans from 97 facilities. The study tested how dose, noise, and resolution are variable among different facilities and reported both the intrafacility and interfacility variabilities. The study further aimed to identify which clinical scan parameters are most related to the observed variability in scan dose, noise, and resolution as a guide toward improving consistency in clinical imaging.

2 Methods

2.1 Study Framework

Scans belonging to (1) chest and (2) combined abdomen-pelvis scan protocols were identified for 97 facilities. On each diagnostic scan series, noise and resolution were measured and combined with dose estimates and Digital Imaging and Communication in Medicine (DICOM) standard information corresponding to the scan acquisition into a dataset. The considered dose metrics were the volumetric computed tomography dose index ($CTDI_{vol}$ in mGy) and the size-specific dose estimate (SSDE in mGy). Statistical models were used to assess: (1) the impact of facilities on image quality, (2) the interfacility variability, (3) the intrafacility variability, and (4) the scan parameter settings with highest impact on the observed results.

2.2 Multi-Center Dataset and Patient Cohort

2.2.1 Statistics on facilities included in the study

Cases for this study were drawn from a retrospective, IRB-exempt, multi-center survey of image quality, dose, and scan parameters of adult scans for 97 facilities that were users of Imalogix dose monitoring system during a 6-month period from June 2018 to January 2019. Data for this study were collected in partnership with the Imalogix Research Institute, which is a conduit to access deidentified DICOM scan information and dose information for all CT scans from participating facilities. All scanners were under the customary imaging physics quality control oversight. Through an agreement with Author Institution Redacted, previously developed algorithms for automated measurements of noise and resolution of CT exams were applied to the cases thus measuring resolution and noise of the images.^{8,9} The Imalogix Research Institute utilized Author Institution Redacted-written, image-quality-measuring, executable code and provided the measurement results along with deidentified dose and DICOM information for review and analysis by the Author Institution Redacted research team, which was blind to information about the participating facilities.

2.2.2 Determination of scan cohort

From all scans in the multi-center survey, two protocols were selected, namely “chest” ($N = 34,137$) and “abdomen-pelvis” ($N = 53,492$) CT exams, as they constitute two most commonly performed CT procedures today.^{13,14} These protocol types were composed of multiple protocol subtypes, which were all included in the study. The breakdown of the frequency of these protocols is shown in Appendix Table 3. In total, 87,629 CT scans were analyzed and contributed to results of this study.

2.3 Image Quality Measurements

2.3.1 Noise measurement

Noise measurements were made according to a previously published methodology.⁹ The method identified regions of soft tissue in the patient's body and measured the standard deviation using a 9 pixel \times 9 pixel square region of interest in the soft tissues. All standard deviation measurements in a particular scan were compiled into a histogram and the mode of that histogram was reported as the noise estimate for the entire scan.

2.3.2 Resolution measurement

Resolution measurements were made using a previously published methodology.⁸ The method used sections of the air-skin interface of the patient as an edge to measure an oversampled edge spread function, which was differentiated and Fourier transformed to yield a modulation transfer function (MTF) at the air-skin interface. Sections of the air-skin interface were binned by their radial distance from scanner isocenter, and an MTF was measured separately for each radial distance bin. The frequency in which the MTF is at 0.50 (f_{50} , mm^{-1}) was calculated for each MTF and measured f_{50} values were interpolated by least squares fit to the f_{50} at a reference distance of 5 cm from isocenter. This reference f_{50} value was reported for all possible cases. To remove uncertainty in the measurement process, suspicious f_{50} values were deemed as owing to bad fits and were discarded leading to a missing rate for the resolution in 65.8% of chest cases and 72.9% of abdomen-pelvis cases.

2.4 Dose Data

The scanner-reported CTDI_{vol} in mGy of each scan series were included in the database. The information was extracted from data provided by the scanner manufacturer, either from the DICOM Structured Dose Report, the DICOM header information, or optical recognition of the "Dose Report Screen" of each exam.

2.5 Relevant DICOM Header Data

DICOM header protocol technique data were acquired and compiled for each series by the Imalox Research Institute to be incorporated into the model. The list of DICOM header tags that were incorporated is shown in Appendix Table 4.

2.6 Statistical Analysis

Statistical analysis was carried out to answer four questions of note regarding the consistency of imaging. The first was whether or not image quality (i.e., noise and resolution) and dose varied as a function of facility. The second was to determine the magnitude of variability in noise, resolution, and dose across patients within each facility ("intrafacility variability"). The third was to estimate the variability in noise, resolution, and dose across different facilities ("interfacility variability"). The fourth was to identify which scan parameters were most significant predictors of image quality and dose performance and thus might be helpful as a path to achieve improved consistency.

2.6.1 Data transformation and prediction intervals

Three pairwise 95% probability prediction intervals were assessed with the noise, resolution, and dose treated as outcomes. The univariate distributions of noise and dose were found to be skewed, so a log transform was applied to these distributions. Resolution values were not found to be skewed, so no transformation was made to them.

2.6.2 Variability at and across facilities: random effects model A

To test the impact of facility on image quality or dose of a scan, a set of linear mixed-effects models was fit to the data for each of the outcomes of noise, resolution, and dose. A different model was fit for each outcome (i.e., noise, resolution, and dose) and protocol subset (i.e., abdomen-pelvis and chest) resulting in a total of six models. Facility ID was incorporated as a random effect, with no other fixed effects. Observations with missing features or outcomes were dropped when fitting the models, so that estimates of a subset of the total 97 facilities were reported. These models were also used to estimate the intrafacility and interfacility variability.

Effect of facility. Type III test of p -value of the facility ID was used, with $p < 0.05$ denoting significance, to assess the role that facilities played in affecting noise, dose, and resolution. If the facility ID was found to be a statistically significant factor, then noise, resolution, and dose were deemed different across these facilities.

Estimate of interfacility variability. To estimate the degree of variability of image quality and dose across multiple facilities, the parameter estimates of the facility ID random intercept term was reported for each of the six outcome-by-protocol models. These magnitudes estimates signify the expected noise, resolution, or dose at each facility. Differences in the estimates reflect systematic differences in image quality and dose between facilities. Facilities that are similar in noise, resolution, or dose will have similar estimates for the corresponding outcome (and vice versa).

Estimate of intrafacility variability. To estimate the degree of variability of image quality and dose across the patients at each given facility, the 95% confidence limit of estimate the facility ID random intercept term was reported for each of the facilities considered. The width of this interval reflects the variability among noise, resolution, and dose for each facility. A narrower confidence interval in the facility ID estimate denotes a facility with greater consistency in the outcome. Gaussian kernel density estimates (KDE) of the bivariate probability density function of noise, resolution, and dose were calculated for abdomen-pelvis and chest exams at every facility (using “Gaussian_KDE” within SciPy’s “stats” module, using Scott’s rule for bandwidth estimation).^{16,17}

2.6.3 Identification of relevant imaging parameters: mixed effects model B

To understand the source of variability, a second set of six linear mixed-effects models was fit to the data. This time, a list of potentially important DICOM tag data corresponding to scan parameters and protocol technique was related to outcomes of noise, resolution, and dose using a mixed-effects model. Features consisted of changeable scan technique parameters (e.g., convolution kernel, tube voltage, tube rotation time, and average tube current) and accounted for differences in patient characteristics (e.g., effective diameter, age, and sex). The mixed-effects models consisted of a combination of fixed and random effects. The full list of features used in the model is shown in Appendix Table 4. Each feature was nested under facility ID random effect to account for systematic differences at different facilities. The allocation of features as fixed or random effects was different for each of the six models and was assigned to result in the best fitting of the models. Table 2 lists the fixed effects used in the fitting of each model. The analogous information regarding the random effects terms are shown in Fig. 5 in Appendix. As before, observations with missing features or outcomes were dropped when fitting the models, so that subsets of the data from the 97 facilities were reported. p -values for type III tests of all features in the model were reported. A list, ranked by the magnitude of each term’s t -value with $p < 0.05$ in each of the six models, was reported for relevant fixed effects.

3 Results

Gaussian KDE of the probability density function of noise, resolution, and dose were calculated for abdomen-pelvis and chest exams. The contour areas where the bivariate kernel density rises above 50% of max are shown in Fig. 2. The results of the random effect model A are shown in Table 1, indicating that the facility ID random intercept term is significant in all six models.

Figure 3 shows the estimates and associated confidence intervals of the facility ID random intercept terms from random effects model A for each of the six models. The standard deviation of the estimates for noise, resolution, and dose for abdomen-pelvis exams across facilities were 0.27, 0.06, and 0.20, respectively. The corresponding values for chest exams were 0.25, 0.07, and 0.23, respectively. The average width of the 95% confidence interval at each facility in noise, resolution, and dose were 0.50, 0.17, and 0.77, respectively, in abdomen-pelvis exams. The corresponding values for chest exams were 1.06, 0.54, and 0.94, respectively. The greatest Pearson’s correlation coefficient between the estimates for noise, resolution, or dose at facilities and the associated confidence interval widths for either abdomen-pelvis or chest exams was 0.34.

The results of a type III test for significant features in mixed effects model B are shown for an abridged list of features in Fig. 4. Figure 4 also details the list of which features were most impactful on noise, resolution, and dose. Fit statistics regarding the random effects and residual error are shown in Fig. 5 in Appendix.

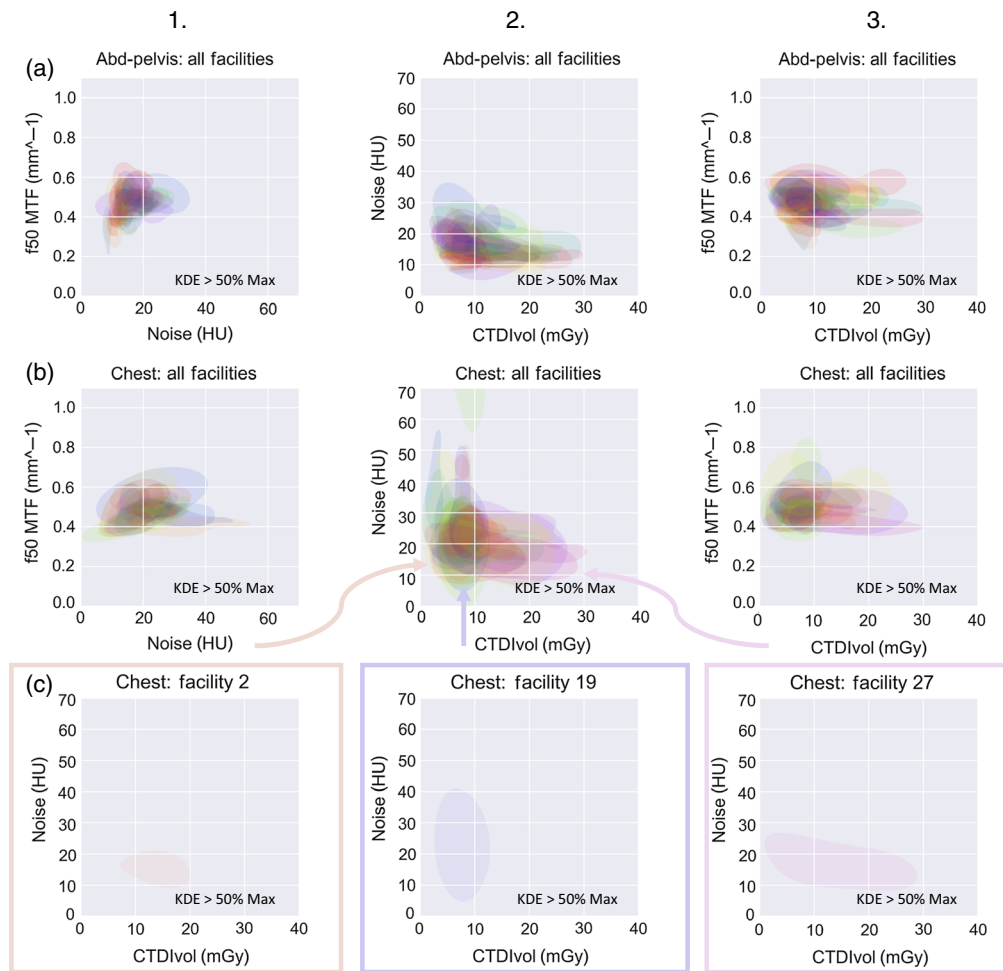


Fig. 2 Relationships between image quality and dose at each facility (differentiated by color) for (a) abdomen-pelvic exams and (b) chest exams. (c) A carve-out of the regions of three sample facilities for the noise versus dose plot [(b), column 2]. In all columns and (a)–(c), each color area represents the range of likely values for a facility via Gaussian KDE. Displayed are the contour areas of 50% of max of the KDE (this used to estimate probability density function) regions for noise, resolution, or dose. Variability in the imaging enterprise shows itself in two ways: (1) inter-facility variability is demonstrated by the different $x - y$ locations of each of the colored regions and (2) higher intrafacility variability is demonstrated by those regions that encompass a greater area (and thus less consistent image quality/dose values). Each of three example facilities in (c) exhibits a different amount of intrafacility variability in image quality and dose (denoted by their areas), as well as interfacility variability between them (denoted by their locations).

Table 1 The fit results of random effect model A show that the noise, resolution, and dose are different at different facilities for both abdomen-pelvis and chest exams.

Outcome	Degrees freedom	Sum of squares	Mean square	F-value	p-value
Abdomen-pelvis					
Log (noise)	94	3828.43	40.73	410.06	<0.0001
Log (dose)	94	2082.23	22.15	93.57	<0.0001
Resolution	92	39.63	0.43	64.13	<0.0001
Chest					
Log (noise)	80	2055.84	25.7	65.94	<0.0001
Log (dose)	80	1870.7	23.38	75.88	<0.0001
Resolution	70	48.16	48.16	41.9	<0.0001

Specifically, the fit results here show that for each of the combinations of outcome and body part, the facility (the random intercept term) has a significant effect. The degrees of freedom denote the number of facilities in the dataset that are considered in each of the models. The missing resolution values are responsible for the lower degrees of freedom in the model. The sum of squares and mean square (sum of squares divided by the degrees of freedom) reflect variability in measurements in the dataset, higher values indicating higher variability. The *F*-value reflects the degree to which the facility-ID is correlated with noise, dose, or resolution. The high *F*-values, and low *p*-values together point to the fact that the noise, dose, and resolution all change significantly at different facilities. Observations with missing covariates or outcomes were dropped when fitting the models so that a subset of 97 facilities were reported.

4 Discussion

Providing consistently high-quality care is the goal of medical practices. For diagnostic imaging, providing high-quality care means providing an accurate diagnosis. Certain image quality conditions make it easier for radiologists to provide confidence in accurate diagnoses. Facilities should aim for diagnostic consistency with requisite radiation dose. In diagnostic imaging, image quality and dose are quantitative metrics that can be used as indicators of consistency in examination performance. Some works have aimed to measure these quantities at and across imaging facilities. This work differs in that it reports on automated assessment of noise and resolution, in addition to dose, as measured from clinical images across a wide pool of facilities.

As in other studies, we see that there are differences between the practice at different institutions and that these choices impact dose and quality.^{1,4,10,11} The facilities exhibited a wide range of noise, resolution, and dose in their examinations—meaning that there are considerably different practices at different facilities. Additionally, there was a considerable range of different variabilities in image quality and dose within facilities. There was, however, little correlation between the magnitude of the image quality and dose estimates at a given facility and their variability at that facility (Pearson correlation coefficient < 0.35 in all cases), as demonstrated for abdomen-pelvis cases in Fig. 3. One may have hypothesized that facilities that image with higher quality (i.e., a lower noise or sharper resolution) would have less variability (higher consistency) in their images. However, that was not proven to be the case. The results suggested that the magnitude or level of image quality or dose at a facility is largely uncorrelated with the consistency in those quantities at the facility.

Next, in an effort to understand this variability in image quality and dose, we ranked the features fit in mixed effects model B. Here we simultaneously account for the effects of many factors on the quality and dose of a scan, including both intentionally selected and changeable scan parameters (e.g., tube kV, mAs, and reconstruction kernel), practical aspects (e.g., patient misalignment), and non-alterable realities (e.g., patient size and scanner model). The *t*-values provided in Fig. 4 give a ranking of the most influential fixed effects on noise, resolution, and dose. These rankings in Table 2 focus only on the magnitude of the effect of that parameter choice value (i.e., a *t*-value of 10 and -10 would be displayed with the same color), but in

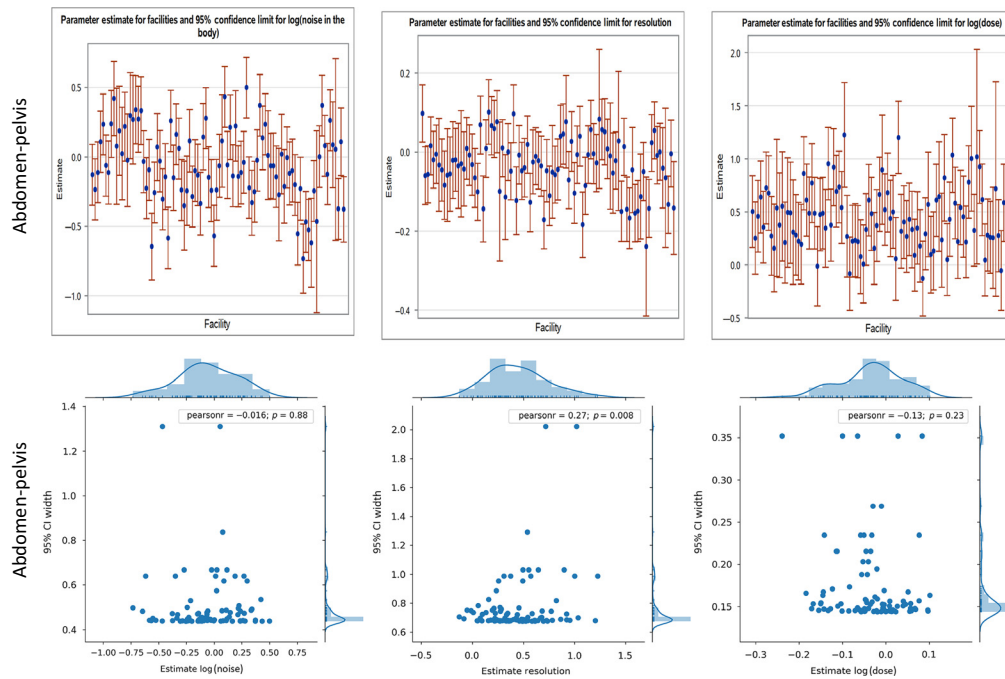


Fig. 3 (a) The best linear unbiased parameter estimates of the facility ID term for random effects model A reflect both high levels of interfacility variability and intrafacility variability (shown in the whisker plots). (b) However, there is little correlation between noise/resolution/dose values at a facility and variability at that facility. The blue points show the most likely value of noise, resolution, and dose at each facility for abdomen-pelvis exams. Thus differences in the location of these points reflect interfacility variability in image quality and dose. The error bars denote the 95% confidence interval of the estimate and reflect intrafacility variability at each facility in the same quantities—i.e., shorter bars at a facility denote more consistency and less variability within the scan population at that facility. (b) shows these estimate values for each facility against the uncertainty in the estimates with associated correlation coefficients between the two values. The low correlation values show that there is little relation between the most likely image quality/dose level at a facility and the variability in quality/dose across patients at that facility. That is, consistency and magnitude of noise/resolution/dose level are not well-correlated.

principal the effect of a parameter choice can increase or decrease image quality or dose. Additionally, although a parameter could have multiple nested values (e.g., scanner model can take on values of “Siemens Flash,” “GE Revolution”, etc., each with their own effect on image quality and dose), the maximum magnitude for each category is shown as a means of understanding broadly which factors have the greatest effect of image quality and dose. The most impactful changeable parameter for noise (for both exam types) was the slice thickness followed by reconstructed field-of-view (FOV); for resolution were the reconstructed FOV and the reconstruction kernel selection; and for dose models were tube current (mAs) and tube voltage (kV). Additionally, simple practical considerations like mis-centering of the patient in the scanner were found to impact the quality and dose, which is consistent with the results of other studies.^{18–20} The impact of scanner technology on quality, which was previously observed for noise, is also present for resolution and dose.⁶ Each of these insights generates potential avenues for improvement of consistency at and across facilities. Although the patients’ sex was found to have a significant effect on noise, resolution, and dose for all abdomen models, and on noise and dose for chest models, the magnitude of the estimates of the patient sex fixed effect was close to zero, and thus the effect of patient sex may be disregarded.

Both of the rankings of features for noise and dose models for abdomen-pelvis and chest were considerably similar. The Spearman correlation coefficient of the magnitude of the t -values for abdomen-pelvis and chest models for noise and dose were 0.76 and 0.86, respectively, (for resolution, the corresponding value was 0.52) suggesting the ranking of the importance of fixed

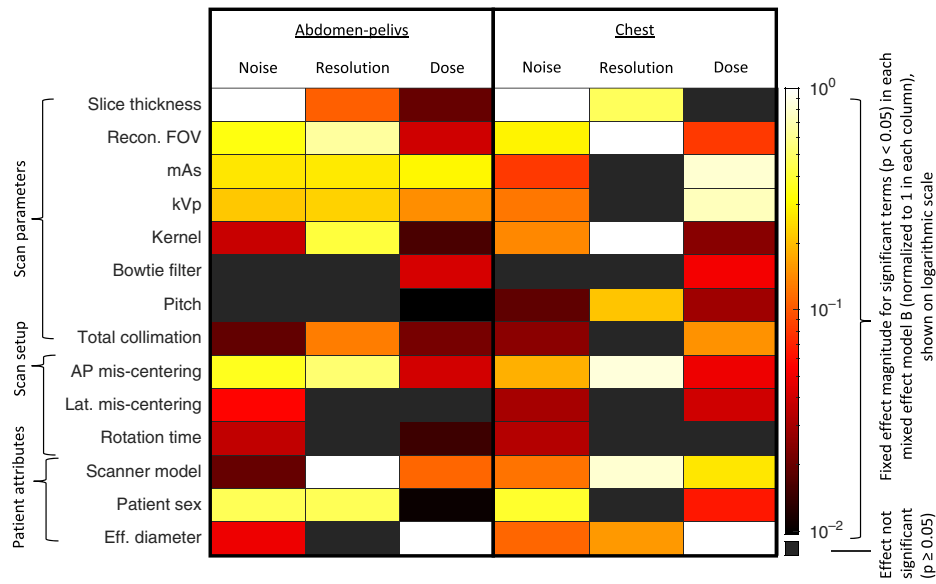


Fig. 4 A summary of the fit results of a mixed effects model B for abdomen-pelvis and chest exams. This figure shows which fixed effects were most explanatory for resolution, noise, and dose (with $p < 0.05$). Colors denote the magnitude of an effect (using log scale). Black cells denote the effects that were not significant. For discrete features with sublevels, sublevel with largest magnitude is shown.

Table 2 A ranked list in descending order of the fixed effects that were found to be significant ($p < 0.05$) in mixed effects model B for abdomen-pelvis and chest exams. For discrete features with sublevels, only the sublevel with largest t -value magnitude is shown. Features that appear toward the top of the list are more impactful on the corresponding outcome.

	Abdomen-pelvis			Chest		
	Noise	Resolution	Dose	Noise	Resolution	Dose
1	Slice thickness	Scanner model	Eff. diameter	Slice thickness	Kernel	Eff. diameter
2	Patient sex	Recon. FOV	mAs	Patient sex	Recon. FOV	mAs
3	AP. mis-centering	AP. mis-centering	kVp	Recon. FOV	AP. mis-centering	kVp
4	Recon. FOV	Patient sex	Scanner model	AP. mis-centering	Scanner model	Scanner model
5	mAs	Kernel	Bowtie filter	Kernel	Slice thickness	Total collimation
6	kVp	mAs	AP. mis-centering	kVp	Pitch	Recon. FOV
7	Lat. mis-centering	kVp	Recon. FOV	Scanner model	Eff. diameter	Patient sex
8	Eff. diameter	Total collimation	Total collimation	Eff. diameter		Bowtie filter
9	Kernel	Slice thickness	Slice thickness	mAs		AP. mis-centering
10	Rotation time		Kernel	Rotation time		Lat. mis-centering
11	Scanner model		Rotation time	Lat. mis-centering		Pitch
12	Total collimation		Patient sex	Total collimation		Kernel
13			Pitch	Pitch		

effects for the abdomen-pelvis and chest models was similar in noise and dose. This similarity in ranking makes sense as the same physics affects both noise and dose in both anatomical regions, but allows for differences in the settings that facilities use to accommodate for necessary trade-offs in imaging for the two anatomical regions. For resolution, concordance between the two exam types was weaker, likely due to differences in the applied protocols for the two regions. This lowered concordance may also be an artifact of a poorer fit of mixed effects model B for the resolution data (fit information displayed in Appendix Table 4). Appendix Table 4 shows the corresponding results for the random effects used to control variability in the models and shows how four of the six instances of mixed effects model B presented excellent fits to the data. The models fit for resolution outcomes had a weaker fit to the data, likely owing to missing values, an issue that needs to be improved in the future.

This study had some limitations. In the process of measuring image quality and dose over such a wide population of facilities and scans, some intentional simplifications to the quantification and methodology were necessary. The metrics of image quality used in this study were task-generic metrics. Although task-generic methods of measuring image quality are appropriate for quantifying the perceptual properties of the image, the ideal measure of quality would be to use a clinically relevant task-specific measure (e.g., detectability index), which is strongly correlated with task performance (e.g., lesion detection^{21–23}). Using a task-specific measure of image quality would allow for a specific investigation of how noise and resolution combine to influence perception and thus diagnosis, which is the ultimate goal. For example, using only task-generic image quality metrics, it may not be clear which of the panels in Fig. 1 shows the optimal trade-off between the resolution and noise for perception of a specified task. High resolution, for example, may not always be deemed optimal, as enhanced sharpness of the pathology of interest may be counter-balanced with enhanced features of or artifacts in the underlying background that hinder the interpretation of the pathology. This points out the need for task-based evaluation and optimization. Other studies have made important strides to investigate interfacility variability via task-based image quality but were limited to the use of phantom images in their analysis.²⁴ As a result of this limitation, phantom-based studies are necessarily limited by the number of different types of “patients” the phantoms can represent. This study used measures of image quality, which were derived entirely from patient images and thus patient-specific. Though task-generic themselves, noise and resolution of f_{50} are relevant measures of image quality and contributors to task-based image performance.

One of the results of this study is a report of the magnitude of variability of scan quality and dose at and across facilities. To do so, we sampled the image quality and dose for 97 different facilities that represented a wide range of facility types and sizes. However, this may still not be a perfectly random sampling of facilities. All facilities considered in this study were clients of the Imalogix dose monitoring system. One might say that these facilities may be preferentially more attuned to the quality and doses that they deliver to their patients, and that the interfacility statistics reported here may not be representative of the wider population of imaging facilities. Thus we suggest that these statistics may be a lower bound on the variability across facilities. Some of the variability in image quality and dose between patients may be conscious and intentional, even necessary, as a result of differences in clinical information or pathology of interest. However, many such differences are not governed by conscious choices. Further, there remain significant differences between facilities. This study makes an effort to control for certain aspects of the imaging process that may contribute to this variability, thus providing an outcome-based guidance to improve the consistency of CT images. The list of the controlling factors, however, was not exhaustive. For example, some aspects of CT equipment technology, such as the use of iterative reconstruction or tube current modulation, were not specifically accounted for. Further studies can increase the number and scope of features and generalizability of the estimation across the national profile.

5 Conclusions

The consistency in image quality and dose and key metrics in radiology practice are reported for 97 facilities. There are considerable differences reported in both the magnitude and consistency

of image quality and dose at and across facilities. The results can be used on a basis to form peer-based criteria for acceptable levels of consistency based on best practices. A ranked list of changeable parameters that affect noise, resolution, and dose for abdomen-pelvis and chest exams is reported. This list can be considered as a first step toward improving consistency in clinical imaging.

6 Appendix: Materials

Table 3 shows the breakdown of which protocol sub-types were included in the study. Table 4 gives the list of predictors used in Mixed Effects Model B and their origins.

Table 3 The frequency of different protocols within the abdomen-pelvis and chest protocol types included in the study. Each of the protocols at each facility was matched to a corresponding RadLex protocol where possible, or to a more specific subprotocol named in the style of the RadLex Playbook.¹⁵

Protocols subtype	Frequency	Percent
Abdomen-pelvis		
Abd pelvis	9338	17.46
Abd pelvis adrn1 W IVCon	2	0.00
Abd pelvis appx	142	0.27
Abd pelvis appx W IVCon	1	0.00
Abd pelvis colongrphy	129	0.24
Abd pelvis entero	144	0.27
Abd pelvis entero multiph	33	0.06
Abd pelvis entero W IVCon	51	0.10
Abd pelvis kidney calc	9878	18.47
Abd pelvis kidney calc WO IVCon	127	0.24
Abd pelvis kidney cancer	16	0.03
Abd pelvis kidney cancer W IVCon	4	0.01
Abd pelvis kidney WO and W IVCon	4	0.01
Abd pelvis lo dose kidney calc	107	0.20
Abd pelvis multiph	254	0.47
Abd pelvis multiph liver	423	0.79
Abd pelvis multiph panc	85	0.16
Abd pelvis multiph WO and W IVCon	1	0.00
Abd pelvis uro	3604	6.74
Abd pelvis uro multiph	45	0.08
Abd pelvis uro W IVCon	240	0.45
Abd pelvis uro WO and W IVCon	56	0.10
Abd pelvis W IVCon	20,206	37.77
Abd pelvis WO and W IVCon	1644	3.07
Abd pelvis WO IVCon	6958	13.01

Table 3 (Continued).

Protocols subtype	Frequency	Percent
Chest		
Chst	7781	22.79
Chst 3D image airway	35	0.10
Chst airway	22	0.06
Chst heart	9	0.03
Chst heart calc score	22	0.06
Chst heart W IVCon	1	0.00
Chst hi res	1518	4.45
Chst hi res WO IVCon	21	0.06
Chst liver multiph	27	0.08
Chst lo dose	57	0.17
Chst lo dose ltd nodule	1	0.00
Chst lo dose screen	2010	5.89
Chst lo dose WO IVCon	274	0.80
Chst ltd nodule WO IVCon	105	0.31
Chst pulm arts embo	1669	4.89
Chst pulm arts embo lo dose	1034	3.03
Chst pulm arts embo lo dose W IVCon	5	0.01
Chst pulm arts embo W IVCon	5768	16.90
Chst pulm arts lo dose WO IVCon	3	0.01
Chst pulm vns	58	0.17
Chst pulm vns W IVCon	2	0.01
Chst research	8	0.02
Chst super dim	878	2.57
Chst super dim W IVCon	18	0.05
Chst W IVCon	2814	8.24
Chst WO and W IVCon	19	0.06
Chst WO IVCon	9951	29.15
RT plan breast	26	0.08
Stlth chst	1	0.00

Table 4 A list of the predictors used in mixed effects model B and their origins.

Variable	Description	Source
Categorical predictors		
T-shirt size	Estimated t-shirt size of the patient	Calculated
Scanner model	Model of scanner	DICOM header
Patient sex	Gender	DICOM header
Scan protocol	Matched protocol of the series	Calculated
Kernel	Reconstruction kernel	DICOM header
Bowtie filter	Bowtie filter type	DICOM header
Intercept (facility)	Facility that performed the exam	Assigned
Continuous predictors		
kVp	Kilo-voltage potential of the x-ray tube	DICOM header
Reconstruction diameter	Diameter of the reconstructed image	DICOM header
Slice thickness	Thickness of reconstructed slice	DICOM header
mAs	Average milli-ampere seconds of the x-ray tube	DICOM header
CTDI*	Radiation output of x-ray tube (*included only in “noise” and “resolution” models)	DICOM Structured Dose Report, or optical recognition of the Dose Report Screen
DLP	Dose-by-scan-length product	DICOM Structured Dose Report, or optical recognition of the Dose Report Screen
Effective size	Effective size of the patient	Calculated
Effective dose	Effective dose to the patient	Calculated
Rotation time	Amount of time of a single revolution of the CT gantry	DICOM header
Max slice thickness	Max thickness of reconstructed slice in acquisition	DICOM header
Total collimation width	The width of the total beam collimation (in mm) over the area of active x-ray detection	DICOM header
Pitch	Pitch of the CT scan	DICOM header
SSDE	Size-specific dose estimate	Calculated
Lateral mis-centering	Offset from isocenter in lateral direction	Calculated
AP mis-centering	Offset from isocenter in anterior-posterior direction	Calculated

Figure 5 contains the fit result information for random effects in Mixed Effects Model B. Table A details how much variance is explained by each random effect. Table B shows the number of facilities with a significant p -value for the random effect.

A: Random effect explained variance							
Abdomen-pelvis				Chest			
Noise				Noise			
Covariates	Estimate	Percentage of Total Variance		Covariates	Estimate	Percentage of Total Variance	
Facility ID	0.71	36.98%		Facility ID	0.08411	15.53%	
Rotation time	0.1	5.20%		Rotation time	0.3093	57.12%	
Pitch	1.047	54.54%		Pitch	0.05589	10.32%	
Kernel	0.046	2.38%		Kernel	0.06079	11.23%	
Residual	0.017	0.89%		Residual	0.03143	5.80%	
Dose				Dose			
Covariates	Estimate	Percentage of total variance		Covariates	Estimate	Percentage of total variance	
Facility ID	0.359	42.00%		Facility ID	0.4083	45.37%	
Rotation time	0.03	3.55%		Rotation time	0.2606	28.95%	
Pitch	0.461	53.89%		Pitch	0.2238	24.87%	
Residual	0.005	0.55%		Residual	0.007325	0.81%	
Resolution				Resolution			
Covariates	Estimate	Percentage of Total Variance		Covariates	Estimate	Percentage of Total Variance	
Facility ID	0.0002	2.09%		Kernel	0.005392	30.80%	
Kernel	0.001	17.21%		Rotation time	0.00591	33.75%	
Residual	0.006	80.70%		Residual	0.006207	35.45%	
B: Random effect significance							
Abdomen-pelvis				Chest			
Random effects				Random effects			
Significant in N facilities				Significant in N facilities			
Feature	Noise	Resolution	Dose	Feature	Noise	Resolution	Dose
Intercept	5		8	Intercept	5		15
Rotation time	14		17	Rotation time	22	5	16
Pitch	9		9	Pitch	8		13
Kernel		14		Kernel	2	9	

Fig. 5 Random effects from mixed effects model B indicating the contribution of leading protocol factors to the total variability in noise/resolution/dose between scans. Random effects with greater contribution are denoted by a darker green. The residual terms account for the fixed effects as well as the estimation error of the models. Table A details how much variance is explained by each random effect. Table B shows the number of facilities with a significant p -value for the indicated protocol factor. A darker green denotes which effects were more commonly significant at facilities.

Disclosures

Donald Frush is chair of the Image Gently Steering Committee. Ehsan Samei lists relationships with GE, Siemens, Bracco, Imallogix, 12Sigma, SunNuclear, Metis Health Analytics, Cambridge University Press, and Wiley & Sons. Imallogix Research Institute provided the service to run the Duke-written executable code and provided the aggregate data to the Duke team, which exercised full control of the code, the data, experimental design, and data analysis.

Acknowledgments

This work was determined to be IRB exempt by the Duke Health Institutional Review Board: IRB #Pro00102344. This project was funded in part by an educational fellowship from Imallogix.

References

1. K. M. Kanal et al., "Variation in pediatric head CT imaging protocols in Washington state," *J. Am. Coll. Radiol.* **8**(4), 242–250 (2011).
2. D. Thickman et al., "National variation of technical factors in computed tomography of the head," *J. Comput. Assist. Tomogr.* **37**(4), 486–492 (2013).

3. D. B. Larson, K. J. Strauss, and D. J. Podberesky, "Toward large-scale process control to enable consistent ct radiation dose optimization," *Am. J. Roentgenol.* **204**(5), 959–966 (2015).
4. J. R. Marin et al., "Variation in pediatric cervical spine computed tomography radiation dose index," *Acad. Emerg. Med.* **22**(12), 1499–1505 (2015).
5. R. G. Abramson, "Variability in radiology practice in the United States: a former teleradiologist's perspective," *Radiology* **263**(2), 318–322 (2012).
6. F. Ria et al., "Expanding the concept of diagnostic reference levels to noise and dose reference levels in CT," *Am. J. Roentgenol.* **213**, 889–894 (2019).
7. Institute of Medicine; National Academies of Sciences, Engineering, and Medicine, *Improving Diagnosis in Health Care*, E. P. Balogh, B. T. Miller, and J. R. Ball, Eds., The National Academies Press, Washington, DC (2015).
8. J. Sanders, L. Hurwitz, and E. Samei, "Patient-specific quantification of image quality: an automated method for measuring spatial resolution in clinical CT images," *Med. Phys.* **43**(10), 5330–5338 (2016).
9. O. Christianson et al., "Automated technique to measure noise in clinical CT examinations," *Am. J. Roentgenol.* **205**(1), W93–W99 (2015).
10. K. M. Kanal et al., "Variation in CT pediatric head examination radiation dose: results from a national survey," *Am. J. Roentgenol.* **204**(3), W293–W301 (2015).
11. K. M. Kanal et al., "U.S. diagnostic reference levels and achievable doses for 10 adult CT examinations," *Radiology* **284**(1), 120–133 (2017).
12. M. A. King et al., "Radiation exposure from pediatric head CT: a bi-institutional study," *Pediatr. Radiol.* **39**(10), 1059–1065 (2009).
13. NCRP, *NCRP Report 184: Medical Radiation Exposure of Patients in the United States* (2019).
14. F. A. Mettler et al., "Patient exposure from radiologic and nuclear medicine procedures in the United States: procedure volume and effective dose for the period 2006–2016," *Radiology* **295**(2), 418–427 (2020).
15. RadLex Radiology Lexicon, "RSNA," <https://www.rsna.org/en/practice-tools/data-tools-and-standards/radlex-radiology-lexicon> (accessed 13 March 2020).
16. P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods* **17**(3), 261–272 (2020).
17. SciPy Community, "SciPy v1.6.0 reference guide," 2020, https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html (accessed 3 Feb. 2021).
18. M. A. Habibzadeh et al., "Impact of miscentering on patient dose and image noise in x-ray CT imaging: phantom and clinical studies," *Phys. Med.* **28**(3), 191–199 (2012).
19. T. Kaasalainen et al., "Effect of patient centering on patient dose and image noise in chest CT," *Am. J. Roentgenol.* **203**(1), 123–130 (2014).
20. T. Toth, Z. Ge, and M. P. Daly, "The influence of patient centering on CT dose and image noise," *Med. Phys.* **34**(7), 3093–3101 (2007).
21. T. B. Smith, J. Solomon, and E. Samei, "Estimating detectability index in vivo: development and validation of an automated methodology," *J. Med. Imaging* **5**(3), 031403 (2017).
22. S. Richard et al., "Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms," *Med. Phys.* **39**(7), 4115–4122 (2012).
23. P. Sharp et al., "Report 54," *J. Int. Comm. Radiat. Units Meas.* **os28**(1), 1–5 (1996).
24. D. Racine et al., "Task-based quantification of image quality using a model observer in abdominal CT: a multicentre study," *Eur. Radiol.* **28**(12), 5203–5210 (2018).

Taylor B. Smith received his doctoral degree in medical physics from Duke University in 2020. His expertise is in patient-specific measures of image quality for x-ray computed tomography.

Shuaiqi Zhang is a biostatistician in Duke Clinical Research Institution. Her research interests is real world evidence, and her work is using various of database and statistical methods such as mix model Implement propensity matching/weighting, regression, LASSO, random forest, survival analysis and landmark analysis to provide number support for medical research.

Alaattin Erkanli is an associated professor of statistics at the Department of Biostatistics and Bioinformatics, Duke University. He has extensive experience in application of Bayesian and classical methods to complex data problems, and in addition to teaching duties, he currently serves as a faculty statistician in the Biostatistics and Epidemiology Research Core supervising several master level statisticians. He served as the leading statistical expert in preparation of this paper.

Donald Frush, MD, is a professor of radiology in the Division of Pediatric Radiology, Duke Medical Center. His research interests are predominantly in pediatric body CT, including technology assessment, image quality, radiation dosimetry and protection, and risk communication in medical imaging. Other areas of investigation include CT applications in children and patient safety in radiology. International organizational affiliations have included the IAEA, WHO, and International Society of Radiology.

Ehsan Samei, PhD, DABR, FAAPM, FSPIE, FAIMBE, FIOMP, FACR, is a tenured professor and chief imaging physicist at Duke University Health System. He has authored over 300 refereed papers, he is passionate about bridging the gap between scientific scholarship and clinical practice through virtual clinical trials and clinically-relevant imaging metrics, such that they could enable optimum quantitative use of imaging, realization of translational research, and clinical processes that are based on evidence.