Evaluation of over 100 scanner-years of computed tomography daily quality control data

Jessica L. Nute

*Department of Imaging Physics, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030
and University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas 77030*

John Rong

Department of Imaging Physics, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030

Donna M. Stevens

Northwest Permanente, PC, Kaiser Sunnyside Medical Center, Clackamas, Oregon 97015

Brandan J. Darensbourg and Jing Cheng

*Department of Imaging Physics Operations, University of Texas M.D. Anderson Cancer Center,
Houston, Texas 77030*

Wei Wei and Brian P. Hobbs

Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030

Dianna D. Cody^{a)}

Department of Imaging Physics, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030

(Received 14 September 2012; revised 8 March 2013; accepted for publication 23 March 2013;
published 19 April 2013)

Purpose: The results of a long-term, comprehensive CT quality control (QC) program were analyzed to investigate differences in failure rates based on QC test, scanner utilization pattern, and number of channels, as well as explore issues regarding testing frequency.

Methods: CT QC data were collected over a 4-yr period for 26 CT scanners representing two different vendors and using three different QC programs culminating in over 100 scanner-years of QC data. QC tests analyzed included water tests [mean CT number, standard deviation, and uniformity], linearity tests [air, water, and acrylic], and artifact analysis [water phantom and large phantom]. The data were organized based on scanner use, number of channels, scanner modality, and QC test. Logistic regression model analysis with generalized estimating equation method was used to estimate failure rates for each group.

Results: A significant difference between failure rates with respect to QC test was found (p -value = 0.02). Large phantom artifacts, standard deviation of water, and water phantom artifacts had the three highest failure rates. No significant difference was found between failure rates organized by scanner use, scanner modality, or number of channels.

Conclusions: Standard deviation of water is the most important quantitative value to collect as part of a daily QC program. Uniformity and linearity tests have relatively low failure rates and, therefore, may not require daily verification. While its failure rates were moderate, daily artifact analysis is suggested due to its potentially high impact on clinical image quality. Weekly or monthly large phantom artifact analysis is encouraged for those sites possessing an appropriate phantom. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4800796>]

Key words: computed tomography, quality control, quality assurance, artifact

I. INTRODUCTION

All facilities that provide clinical CT scanning services should employ some form of quality control (QC) program to ensure the continued satisfactory operating condition of imaging equipment. For the purposes of this paper, QC will refer to the simple daily tests performed on CT scanners to monitor a select number of basic aspects of scanner performance. Tests performed on an annual basis or as part of acceptance testing or dosimetry measurements, although very important, are outside the scope of this paper.

A strong QC program is important to ensure the overall image quality of clinical CT exams. While there is the poten-

tial that newer CT technology will be more stable and better constructed and, therefore, require less QC, the increase in complexity that accompanies these advances and the adoption of a more quantitative approach within CT applications both suggest a need for more vigilant QC. Maintaining consistent and satisfactory operating conditions of these scanners may reduce repeat clinical scans that would result in increased patient dose, issues with reimbursement limitations, and wasted staff and patient time.¹

CT manufacturers provide a list of scanner-specific tests and tolerances that should nominally be performed on a routine basis. These vendor-provided descriptions can be used as the basis of a structured QC program. In a few cases, the CT

QC program is regulated by state or local authorities. Some states, including New York, New Jersey, and Minnesota, have specific requirements and criteria for performing routine CT QC testing; however, most states have no form of standardized or regulated CT QC testing.²⁻⁵

In the past, several recommendations have been published regarding the specific tests and limit criteria to be included in a CT QC program.⁶⁻¹² These recommendations were not specific to scanner type or vendor. In general, these recommendations included tests for the mean CT number of water, image noise (standard deviation of the CT number of water), and a measure of field uniformity; however, the limits and recommended frequency of these tests varied widely. Recently, the American College of Radiology (ACR) CT accreditation program released its recommendations on a standardized CT QC program.¹³ The recommendations include daily monitoring of mean CT number of water, standard deviation of the CT number of water, and artifact analysis.

The goal of any QC program is to include sufficient testing to ensure good image quality and patient safety but not to include so many tests that it becomes impractical in a clinical environment and unreasonable to the staff. With access to years of archived QC results from a large number of CT systems, the authors noted a unique opportunity to investigate failure rates of specific tests, as well as how utilization pattern and number of channels may affect those failure rates. Analysis of such a large database of QC results could help identify which parameters fail frequently enough to be included as part of a daily QC program and which parameters fail infrequently enough to justify excluding them. In addition, predictive base failure rates could be calculated that would indicate how often, on average, a CT scanner will experience a failed test result once a rigorous QC program has been established. Therefore, a database of CT QC results from a stable QC program was collected over a 4-yr period and was analyzed specifically

to calculate and compare failure rates for those tests included in a typical CT QC program and to discuss potential impacts on testing frequency. We hope that these results will serve as a useful tool for facilities aiming to develop a CT QC program within the framework recommended by the ACR CT accreditation program.

II. MATERIALS AND METHODS

II.A. Extent of data

The archived results of daily QC were collected for 26 scanners from June 2007 to June 2011 (Table I). Of the 26 scanners included in this study, 16 were stand-alone CT systems, five were SPECT/CT systems, and five were PET/CT systems. Eight of the scanners were manufactured by Siemens and 18 by General Electric. The data covered three different QC programs: General Electric Quality Control (GE-QC), Siemens Quality Control (S-QC), and an In-House Quality Control (IH-QC). Over the interval from June 2007 to June 2011, QC results were collected for 858 scanner-months of GE-QC, 364 scanner-months of S-QC, and 902 scanner-months of IH-QC, yielding over 100 scanner-years of QC data. In addition, a large phantom artifact program was implemented in November 2010 and performed weekly on all GE CT scanners, resulting in 99 scanner-months of large phantom artifact scan data over our study interval.

II.B. QC programs and tests

GE-QC was a vendor-specific QC program designed by GE and performed on all GE scanners. The QC phantom consisted of a 22-cm diameter, water-filled acrylic phantom provided by the vendor. The techniques for the daily QC are detailed in Table II for GE scanner models included in this

TABLE I. List of scanners included in the study organized by scanner model and the number of scanner-months' each QC protocol was run on each scanner model. QC programs: General Electric Quality Control (GE-QC), In-House Quality Control (IH-QC), and Siemens Quality Control (S-QC).

Mfr.	Modality	Scanner model	No. of scanners	GE-QC (months)	IH-QC (months)	S-QC (months)	Large phantom (months)
GE	PET/CT	DRX	1	49	49		
		DST ^a	2	76	76		
		DSTE ^a	2	69	69		
		DVCT	1	40	40		
GE	CT	LS+ ^a	4	109	152		16
		LS-16	7	337	338		51
		VCT ^a	4	178	178		32
Siemens	SPECT/CT	Emotion 6 ^a	5			187	
		Emotion 16 ^a	1			44	
Siemens	CT	Sensation 16	1			46	
		Sensation 64	1			44	
		Sensation Open	1			43	
Total months				858	902	364	99
Total years				71.5	75.2	30.3	8.3

^aNote: Over the study interval, one DST scanner was upgraded to a DSTE scanner, two LS+ scanners were upgraded to VCT scanners, and one Emotion 6 scanner was upgraded to an Emotion 16 scanner.

TABLE II. GE-QC program scan parameters (for all scanner models: DFOV = 25 cm, scan mode = helical, kVp = 120).

Scanner model	No. of images	Im Thk (mm)	Pitch	Det Config (mm)	(mA)	Rot Time (s)	(mAs)	Eff (mAs)	SFOV
LS+	1	10	0.75	4×5	190	1	190	253	Head
DST	1	10	0.625	4×2.5	400	1	400	400	Head
LS-16/DRX/DSTE	1	10	0.625	8×2.5	160	1	160	256	Small
VCT/DVCT	1	5	0.51563	8×5	335	0.4	134	260	Head

study. While GE recommends daily tests covering high contrast resolution, low contrast detectability, contrast scale, slice thickness, laser accuracy, noise, and uniformity, our institution limited the daily tests to noise, uniformity, mean water CT number, and standard deviation of water. An in-house automatic QC (AutoQC) software program was developed to enable archival of QC test data from vendor-specific QC programs. The software program was designed to mimic the procedure outlined by the vendor QC program. As such, the types of images used and the analysis performed may differ between GE-QC, S-QC, and the IH-QC.

As part of the GE-specific QC program, a single helical image was acquired in the water-filled portion of the phantom; this image was then transferred to an off-line computer for quantitative analysis by the in-house AutoQC program. The software automatically added three square $20 \times 20 \text{ mm}^2$ regions of interest (ROIs): one in the center of the image and two around the periphery at the 12 and 3 o'clock positions, as per the GE-specific QC procedure (Fig. 1). These ROIs were used to perform the three water tests: mean water CT number, standard deviation of water CT number, and uniformity of water CT number. The mean and standard deviation were measured by calculating the mean pixel value and standard deviation within the central ROI. To calculate the uniformity, the difference in mean pixel value between each of the peripheral ROIs and the central ROI was calculated. The uniformity was then defined as the maximum difference in mean pixel values in the edge ROIs versus the center ROI. The values cal-

culated for these three water tests were then checked against tolerances provided by GE, and the results were posted to an internal website for review. Current tolerances for GE scanner types included in this study are listed in Table III.

The S-QC was a vendor-specific daily constancy check designed by Siemens and performed on all Siemens scanners. The S-QC phantom consisted of several sections, including a water-filled acrylic portion, provided by the vendor. The techniques for the daily QC are detailed in Table IV for Siemens scanner types included in this study. Daily QC recommended by Siemens included mean CT number of water and standard deviation of water. A single sequential image was acquired in the water-filled portion of the phantom; this image was then transferred to an off-line computer for quantitative analysis by the AutoQC software. The software automatically added five square $20 \times 20 \text{ mm}^2$ ROIs to the image: one in the center of the image and four peripheral ROIs at the cardinal directions, as per the Siemens-specific QC procedure. These ROIs were used to calculate the mean water CT number and the uniformity of water CT number. Because five ROIs were applied to the image, the uniformity value was based on four peripheral ROIs rather than the two applied in the GE-QC AutoQC procedure. To allow Siemens standard deviation tolerances to be applied to the results of the AutoQC software's analysis, the standard deviation had to be measured on a different image. As part of the S-QC program, two consecutive water images were automatically subtracted on the scanner to produce a "subtraction image" (Table IV). This subtraction image was identified and transferred to the AutoQC software where a $68 \times 68 \text{ mm}^2$ ROI was drawn on the image and the standard deviation was calculated. The measured standard deviation is then divided by the square root of two to compensate for the use of the subtraction image and to obtain a measurement of noise.¹⁴⁻¹⁶ This modified determination of noise as well as the results of the mean water CT number and uniformity of water CT number were then checked against tolerances provided

FIG. 1. Placement and size of the ROIs applied by the AutoQC software for GE-QC program water tests.

TABLE III. GE-QC program water test tolerances.

Scanner model	Mean CT# water (HU)	Std dev water (HU)	Unif max diff (HU)
LS+	± 3	2.6–3.4	± 3
DST	± 3	2.6–3.4	± 3
LS-16/DRX/DSTE	± 3	2.9–3.5	± 3
VCT/DVCT	± 3	3.8–4.8	± 3

TABLE IV. S-QC program scan parameters (for all scanner models: scan mode = sequential, DFOV = 25 cm).

Scanner model	Test	No. of images	(kV)	Rot time (s)	(mA)	(mAs)	Im Thk (mm)	Det config (mm)
Sensation open	Mean, Unif, artifact	6	120	0.5	250	125	4.8	6 × 4.8
	Std Dev, artifact	2	—	Subtraction image ^a		—	4.8	6 × 4.8
	Artifact	6	140	0.5	200	100	4.8	6 × 4.8
	Artifact	2	—	Subtraction image ^a		—	4.8	6 × 4.8
Sensation 16	Mean, Unif, artifact	2	120	1	250	250	9	2 × 9.0
	Std Dev, artifact	2	—	Subtraction image ^a		—	9	2 × 9.0
	Artifact	3	140	0.75	200	150	3	3 × 3.0
	Artifact	1	—	Subtraction image ^a		—	3	3 × 3.0
Sensation 64	Mean, Unif, artifact	6	120	0.5	250	125	4.8	6 × 4.8
	Std Dev, artifact	2	—	Subtraction image ^a		—	4.8	6 × 4.8
	Artifact	6	140	0.5	200	100	4.8	6 × 4.8
	Artifact	24	—	Subtraction image ^a		—	4.8	6 × 4.8
Emotion 6	Artifact	3	80	1	140	140	4	3 × 4.0
	Artifact	3	110	1	150	150	4	3 × 4.0
	Artifact	1	—	Subtraction image ^a		—	4	3 × 4.0
	Mean, Unif, artifact	3	130	1	200	200	4	3 × 4.0
Emotion 16	Std Dev, artifact	1	—	Subtraction image ^a		—	4	3 × 4.0
	Artifact	3	80	1	140	140	4.8	3 × 4.8
	Artifact	3	110	1	150	150	4.8	3 × 4.8
	Artifact	1	—	Subtraction image ^a		—	4.8	3 × 4.8
	Mean, Unif, artifact	3	130	1	200	200	4.8	3 × 4.8
	Std Dev, artifact	1	—	Subtraction image ^a		—	4.8	3 × 4.8

^aNote: Subtraction image created automatically by subtracting two consecutive water images from the preceding parameter group.

by Siemens and the results were posted to an internal website for review. Current tolerances for Siemens scanner types included in this study are listed in Table V.

While additional sections can be purchased from Siemens to include an area for linearity tests, the Siemens constancy program does not allow the modification necessary to include these tests as part of the daily QC program. Regardless of the outcome of the Siemens constancy check on the scanner, all images from the S-QC program were also forwarded to a physics technologist for artifact analysis. The images were displayed using a narrow grey level setting (window width: 100 HU, window center: 0 HU) and viewed in stack mode (viewer format displaying images individually but allowing the technologist to quickly page up and down through multiple images). Any artifacts discovered in the images were then recorded in an online artifact log and triaged for service if deemed necessary by an experienced physics technologist or physicist.

IH-QC was an in-house QC program that assessed scanner linearity and the presence of artifacts and was performed

solely on GE scanners. The program was designed to supplement the GE-QC program by adding both water phantom artifact analysis and linearity checks to the water tests already performed. As physicians increasingly use CT images more quantitatively, ensuring the linearity of the scanner's attenuation response in addition to the mean water CT number has become more important. Table VI lists the techniques used in the IH-QC program for the scanner types included in this study. Artifact images were gathered using the water-filled portion of the GE vendor-supplied phantom; while linearity scans were performed using the section of the phantom containing an acrylic block typically used for image thickness and spatial resolution determinations. Artifact analysis was

TABLE VI. IH-QC program scan parameters (GE scanners). (For all scanner models: DFOV = 25 cm, SFOV = head, kVp = 120, rotation time = 1 s, scan mode = axial.)

Scanner model	Test	No. of image	Im Thk (mm)	(mA)	Det config (mm)
LS+	Linearity	1	5	240	4 × 1.25
	Artifact	4	1.25	240	4 × 1.25
	Artifact	4	5	240	4 × 5.0
DST	Linearity	1	5	300	4 × 1.25
	Artifact	8	1.25	400	8 × 1.25
	Artifact	8	2.5	400	8 × 2.5
LS-16/DRX/DSTE	Linearity	1	5	240	4 × 1.25
	Artifact	16	0.625	240	16 × 0.625
	Artifact	16	1.25	240	16 × 1.25
VCT/DVCT	Linearity	1	5	240	8 × 0.625
	Artifact	32	0.625	240	32 × 0.625
	Artifact	16	2.5	240	16 × 2.5

TABLE V. S-QC program water test tolerances.

Scanner model	Mean CT# water (HU)	Std dev water (HU)	Unif max diff (HU)
Sensation open	±4.0	9.72–11.88	±4.0
Sensation 16	±4.0	6.35–7.76	±4.0
Sensation 64	±4.0	10.89–13.31	±4.0
Emotion 6	±4.0	10.8–13.2	±4.0
Emotion 16	±4.0	9.27–7.76	±4.0

^aNote: Standard deviation of water derived from subtraction image.

performed on two different image sets. The first was acquired in axial scan mode using a technique that generates the narrowest image thickness possible to detect any issues to a specific detector row. The second image set was acquired in axial scan mode using a technique that calls for the narrowest image thickness possible while still covering the whole detector surface along the z-axis resulting in thicker images than the first set. These images were made available to a physics technologist for visual assessment. The images were viewed in stack mode using a narrow grey level setting (window width: 100 HU, window center: 0 HU). Any artifacts discovered in the images were then recorded in the online artifact log and triaged for service if deemed necessary by an experienced physics technologist or physicist. The images captured for linearity tests were made available to the AutoQC software where three rectangular ROIs were added to the image. The first was a 10×10 pixel ROI in the acrylic insert captured in the image, the second was a 10×10 pixel ROI in the water portion of the phantom outside the insert, and the third was 10×30 pixels and placed in the air outside the phantom but within the display field of view (DFOV). The correct placement of the air ROI was validated by adjusting the window width and level to allow for direct visualization of the edge of the phantom and the edge of the image space. Incorrect placement of the air ROI outside the image space may not be readily apparent and would result in erroneous results. Mean CT numbers for acrylic, water, and air were compared with tolerances established in-house (air: -1000 to -900 HU, water: ± 5 HU, acrylic: 108 – 132 HU), and the results were posted to an internal website for review.

Up to this point, all QC scans had been acquired with a phantom roughly 20 cm in diameter, which assessed a limited portion of the available 50-cm diameter image space. Artifacts can appear outside the central region of the scan field of view (SFOV). Therefore, in November 2010 a large phantom program was added as a weekly QC process for all GE CT scanners. The large phantom was a cylindrical, polyethylene slab with a diameter of 48 cm and a thickness of 5 cm. (This phantom was provided by the vendor for tests performed during the service calibration.) The scan technique included a range of rotation times, mA, and detector configurations to better cover the range of imaging parameters used clinically (Table VII). The set of images covering the extent of the detector and the range of imaging parameters in the protocol are placed in stack mode and reviewed by a physics technologist who checked each image manually for artifacts using a window width of 400 and a level of -85 . If any artifacts were found in the image, the large phantom scan was recorded as having failed and the scanner triaged for service. The type of artifact, its position, and its associated imaging parameters were recorded for each artifact in the acquisition set for later analysis.

II.C. Data analysis

The results of the AutoQC software for the three QC programs (GE-QC, S-QC, and IH-QC) were archived in a database. Irrelevant failure results, such as multiple failed

TABLE VII. Large phantom artifact program scan parameters (routine scanners only). For all scanner models: Scan mode = axial, kVp = 120, SFOV = large, DFOV = 50 cm.

Scanner model	No. of images	Im Thk (mm)	Det. config. (mm)	(mA)	Rot. time (s)	(mAs)
LS+	4	1.25	4×1.25	440	0.8	352
	4	1.25	4×1.25	350	1	350
	4	5	4×5	440	0.8	352
	4	5	4×5	350	1	350
LS-16	16	0.625	16×0.625	440	0.5	220
	16	0.625	16×0.625	365	0.6	219
	16	0.625	16×0.625	315	0.7	221
	16	0.625	16×0.625	275	0.8	220
	16	0.625	16×0.625	220	1	220
	16	1.25	16×1.25	440	0.5	220
	16	1.25	16×1.25	365	0.6	219
	16	1.25	16×1.25	315	0.7	221
	16	1.25	16×1.25	275	0.8	220
	16	1.25	16×1.25	220	1	220
	16	1.25	16×1.25	220	1	220
	16	1.25	16×1.25	220	1	220
VCT	32	0.625	32×0.625	780	0.4	312
	32	0.625	32×0.625	640	0.5	320
	32	0.625	32×0.625	535	0.6	321
	32	0.625	32×0.625	460	0.7	322
	32	0.625	32×0.625	400	0.8	320
	32	0.625	32×0.625	320	1	320
	16	2.5	16×2.5	780	0.4	312
	16	2.5	16×2.5	400	0.8	320
	16	2.5	16×2.5	400	0.8	320
	16	2.5	16×2.5	400	0.8	320

morning scans or failure results due to phantom mispositioning that were repeated correctly with passing results, were removed from the QC result database. The remaining failed QC scan results were then divided by test so that a single QC scan that resulted in more than one test failure (for example, both a failed mean water CT number and a failed standard deviation value) was counted in all appropriate test categories. Test failures were then investigated on a scanner-by-scanner basis. Scanners with repeated test failures over multiple days were investigated for any significant service that might explain the preceding string of test failures. If the failures correlated with a period between a service request and a service date, the string of failures was considered to be due to a single recurring problem, and all but the first failure event were removed from analysis. The final list of remaining test failures was organized by individual scanner and archived. Pass results were similarly organized and reserved for later comparison to failure results.

For each individual scanner, all relevant tests across each relevant QC program were listed with their failure rate calculated from the number of failures divided by the associated total number of days daily QC was performed over the 4-yr study interval. As large phantom scans were performed on a weekly basis, the associated failure rate was calculated using the total number of failures divided by the number of large phantom weekly scans. Because the large phantom program is more extensive, the imaging parameters used in each artifact-containing scan were retained and recorded with the artifact for further analysis.

Scanners, and their associated failure rates, were then organized into groups and labeled based on factors that were thought to potentially influence the characteristics and stability of the machine. First, scanners were organized and labeled by modality based on the idea that the CT components may vary between dedicated CT systems, PET/CT systems, and SPECT/CT systems. Second, scanners were organized and labeled by number of channels to determine if newer technology resulted in higher QC performance. Finally, considering that failure rates may be influenced by scanners' throughput and tube heat load pattern, scanners were organized and labeled based on use. Scanners that performed routine tasks at high throughput were listed as routine scanners and consisted of only GE CT scanners. Scanners that performed interventional scans at an average throughput were listed as interventional scanners and consisted of only Siemens CT scanners. Finally, scanners for which the CT component was part of a hybrid system and had relatively low throughput were listed as hybrid scanners and consisted of both GE PET/CT and Siemens SPECT/CT scanners.

A logistic regression model with generalized estimating equation (GEE) method¹⁷ was used to estimate failure rate by groups defined by QC test, use, and number of channels. The GEE method took into account the correlation between measurements from the same CT machine. The estimated failure rates with corresponding 95% confidence intervals were reported. A score test based on the GEE model was used to assess if there was significant difference between types of failures and groups of machines with respect to failure rate. Due to the limited number of unique CT machines tested, no interaction term was tested in the multivariate final model. Since modality groupings were confounded with use type, a separate univariate logistic regression model with GEE method was used to analyze the modality group data. The estimated failure rates with corresponding 95% confidence intervals were reported. All tests were two-sided, and p -values of 0.05 or less were considered statistically significant. Statistical analysis was carried out using SAS version 9.3 (SAS Institute, Cary, NC). Plotting was performed using Spotfire S+ 8.2 (TIBCO Inc., Somerville, MA).

Any specific scanner model with an abnormally high failure rate was further investigated by producing a histogram of all available results to visualize the distribution of test values.

To visualize long-term trends in individual routine scanners, the average and standard deviation of results from passing QC for mean water CT number and standard deviation of water on a given scanner were graphed month-by-month over the time frame of the study. The inclusion of failure data in the analysis would drastically affect the average QC value for the month and hinder the identification of trends, so therefore the analysis was limited to passing QC values.

Large phantom scan failures organized by number of channels available on scanner were further binned by the rotation time recorded for each large phantom artifact and the results were graphed to determine if large phantom artifact presence is dependent on rotation time.

III. RESULTS

Table VIII(a) shows the results of the multivariate logistic regression model and corresponding 95% confidence intervals for data organized by QC test, use, and number of channels. A significant difference was found between failure rates based on QC test (p -value = 0.02). Large phantom artifacts had the highest failure rate followed by standard deviation of water and water phantom artifacts. No significant difference in failure rates was found based on number of channels (p -value = 0.07) or use type (p -value = 0.64). Table VIII(b) shows the results of the univariate logistical regression model and corresponding 95% confidence intervals for data organized by scanner modality. No significant difference in failure rate based on modality was found (p -value = 0.24). Figures 2–5 summarize these results.

Unusually high failure rates were noted in 64 channel scanners. Based on the estimated 95% confidence intervals, these scanners had significantly higher failure rates than 4-, 8-, and 16-channel scanners. On further investigation, the 64-channel scanner group's higher failure rate was primarily due to failures in standard deviation of water on the 64-channel GE scanners. Constructing a histogram of standard deviation values for all 64-channel GE scanners showed the distribution of values was not centered in the vendor-determined pass range. Failure values comprised the tail of the distribution that crosses the upper tolerance limit (Fig. 6).

Analysis of standard deviation passing results by scanner over the 4-yr period included in this analysis showed variability but no overall trends (data not shown). Analysis of mean water CT number passing results by scanner over time for the 16-channel GE CT scanners ($n = 7$) showed steady decline at an approximate rate of 0.11 HU per month with an abrupt increase in mean water CT number every 12–18 months (Fig. 7).

Analysis of the large phantom artifact data by rotation time showed that decreasing rotation time correlated with a higher number of large phantom artifacts in both 16- and 64-channel scanners (Fig. 8).

FIG. 2. Results of the multivariate logistic regression model for failure rates grouped by scanner use type. The number of scanners included in each group is noted below the group label.

TABLE VIII. Summary of (a) multivariate logistic regression model with GEE method results for failure type, number of channels, and use type and (b) univariate logistic regression model with GEE method results for modality.

		Estimated failure rate (#failures/1000 scan days)	95% lower confidence level	95% upper confidence level	<i>p</i> -value
(a) Multivariate logistic regression model					
Use type	Interventional	0.98	0.39	2.42	0.64
	Hybrid	1.65	0.96	2.84	
	Routine	1.71	1.06	2.74	
Failure type	Water, standard deviation	5.11	3.29	7.93	0.02
	Water, mean CT no.	0.85	0.49	1.47	
	Water, uniformity	0.41	0.23	0.74	
	Linearity, air	0.55	0.22	1.40	
	Linearity, water	0.65	0.29	1.43	
	Linearity, acrylic	1.23	0.55	2.79	
	Water phantom artifact	2.85	1.76	4.60	
	Large phantom artifact	6.63	2.62	16.74	
No. of channel	4	0.74	0.31	1.79	0.07
	6	0.28	0.07	1.06	
	8	2.69	0.93	7.80	
	16	1.33	0.83	2.14	
	24	2.92	1.54	5.52	
	64	3.54	2.47	5.08	
(b) Univariate logistic regression model					
Modality	SPECT/CT	1.20	0.41	3.55	0.24
	PET/CT	2.85	1.31	6.21	
	CT	3.00	2.28	3.94	

IV. DISCUSSION

Of the QC tests investigated in this study, large phantom artifacts had the highest failure rate. The high standard deviation is a result of both the limited sample size for the weekly test as well as high variation in failure rate across individual scanners. The large phantom artifact failure rate for 16-channel scanners was higher than that for 64-channel scanners most likely due to more advanced electronics in the newer 64-channel models as well as the relatively advanced age of

the 16-channel scanners (Fig. 8). The increase in large phantom artifacts at shorter rotation times (faster rotation speed) is consistent with increased speed causing higher levels of mechanical stress on the system and leading to poor connections between electronic components of the data acquisition system. This increase in large phantom artifacts at shorter rotation times is of particular clinical importance as these shorter rotation times are used more often than longer rotation times in clinical practice. While this data may support the testing

FIG. 3. Results of the multivariate logistic regression model for failure rates grouped by QC test. The number of scanners included in each group is noted below the group label.

FIG. 4. Results of the multivariate logistic regression model for QC test failure rates grouped by the scanner's number of channels. The number of scanners included in each group is noted below the group label.

FIG. 5. Results of the univariate logistic regression model for failure rates grouped by scanner modality. The number of scanners included in each group is noted below the group label.

of shorter rotation times exclusively, artifact analysis over a wide range of rotation times is still recommended for the initial phase of large phantom scan implementation. Initial artifact analysis using a range of rotation times would allow for individual sites to determine the effect of rotation time on failure rate for their unique scanner population and modify their large phantom protocol accordingly.

Because most of the artifacts observed on the large phantom occurred outside the active diameter of the daily QC phantom, these artifacts would not be found by a typical 20-cm diameter water phantom-based QC program (Fig. 9). Inclusion of a large phantom scan on a weekly or biweekly basis would help identify and resolve the issues responsible before the artifacts affect the image quality of clinical scans.

Standard deviation of water failures was the highest of any of the daily QC tests and was found to have a significantly higher failure rate than all linearity tests as well as water uniformity and mean CT number. While standard deviation of water failures can be difficult to resolve, even with the help

FIG. 7. Average GE-QC mean water passing values for LS16 Routine GE CT scanners plotted over time.

of a trained CT engineer, these failures have potentially serious clinical implications including failure to identify low contrast lesions as well as repeat exam requests due to poor image quality, resulting in increased patient dose. Assessment of standard deviation of water is, therefore, strongly suggested as a part of a strong daily QC program. While failure rates for mean CT number of water were low, the increase in use of quantitative imaging as well as the simultaneous availability of the mean and standard deviation measures merits the inclusion of mean CT number in daily QC analysis.

Water phantom artifact analysis, while having lower failure rates than both standard deviation of water and large phantom artifact rates, also has a potentially large impact on clinical image quality and potential for mimicking disease.¹⁸ These artifacts are often easy to resolve by repeating the air calibration scan. While some may argue that quantitative analysis of uniformity and noise are sufficient to identify images with inconsistencies, our experience has shown that even those images with severe artifact can result in passing quantitative values. For example, Fig. 10 shows two artifacts found in QC images of Siemens scanners that had passed the constancy check. Thus, for its high impact on image quality, ease of

FIG. 6. Histogram of passing and failure values for standard deviation of 64-channel GE CT (VCT) and PET/CT (DVCT) scanners (n = 5). Passing values are marked in gray, while failure values are marked in black. Test tolerances are represented by black vertical lines.

FIG. 8. Large phantom artifacts for routine GE CT scanners organized by channel number and plotted by rotation time. (Note: No large phantom artifacts were found in images from 4-channel GE CT scanners.)

resolution, and increased sensitivity compared to quantitative analysis, water phantom artifact analysis by a trained CT technologist is suggested as part of a comprehensive daily QC program.

As shown in the histogram of the water standard deviation data for the 64-channel scanners (Fig. 6), the failure values were within a normal distribution of total results and were not outliers. This implies that the high failure rate was due to the inherent variation in the system performance and that the vendor-specified tolerances for the system were inappropriate. Comparing the 64-channel GE scanners to their 4-channel and 16-channel counterparts, differences can be seen in the GE-QC program scan parameters. While 4- and 16-channel scanners used 10-mm thick images to calculate the standard deviation, 64-channel scanners are unable to produce 10-mm thick images and, therefore, the image thickness was limited to 5 mm. Halving the image thickness without altering the other acquisition parameters would increase the standard deviation value in the image by around 40%, which was not fully accounted for in the higher 64-channel standard deviation tolerance applied by GE. On the basis of the range of values shown in Fig. 6, the tolerance for standard deviation measurements in 64-channel GE Routine scanners would more appro-

FIG. 10. Artifacts identified in S-QC images after the scanner had passed its daily constancy check and all water tests. (Left: Small, high intensity, centrally located artifact from a Siemens Sensation 16 CT Scanner. Right: Multiple, blurry, concentric ring artifacts consistent with detector imbalance from a Siemens Emotion 16 SPECT/CT scanner.)

priately be placed at 4–5 HU rather than the current vendor-supplied tolerance of 3.8–4.8 HU. Analysis of histograms, in conjunction with trend analysis, of both pass and failure values for a given test is encouraged for any scanner experiencing higher than normal failure rates. The analysis can aid with evaluation and/or adjustment of tolerance placement and the addition of appropriate warning levels.

Graphs of mean water CT number pass values over time for individual scanners showed steady decline within the pass range, ending with a abrupt increase in mean value. When the dates of the abrupt increases in mean water CT number were compared with service records on the scanners, almost all increases corresponded to x-ray tube replacements. This would indicate that the steady decline in mean value is a reflection of x-ray tube aging, perhaps indicative of an increase in anode material deposition on the inside of the x-ray tube, effectively adding inherent filtration and hardening the beam as a result. Monitoring long-term trends in standard deviation and mean water CT numbers would be valuable in identifying and resolving recurring or degenerative problems, as well as possibly permitting scheduling of tube replacements which could minimize clinic down-time.

FIG. 9. Left: Normal large phantom artifact scan. Center: Large phantom scan with artifact. Right: Large phantom artifact with overlay displaying the relative size of the 20-cm daily QC phantom to the 48-cm large phantom.

As failure rates between modalities, use types, and number of channels were all found to be statistically insignificant, there is no evidence to suggest that subset-specific QC programs should be pursued. This could potentially simplify QC programs at larger institutions with a large range of scanner types. In addition, our suggestion of limiting quantitative analysis to standard deviation of water and mean water CT number is consistent with the major portion of vendor-recommended daily QC.

IV.A. Limitations

Although this study covers over 100 scanner-years of data, the results are limited by the relatively small number of scanners included. These scanners also address only two different vendors: General Electric and Siemens. The results are further influenced by the scanner population at the institution, which is dominated by GE LS16 CT scanners in the Routine category and Emotion 6 SPECT/CT scanners in the Hybrid category. Due to the automated imaging analysis and QC program, it was feasible to collect the linearity data daily on the GE scanners for use in troubleshooting the daily QC data, if needed, and to verify the long-term stability of these parameters. Unfortunately, no equivalent addition could be made to the vendor-supplied QC on Siemens scanners due to the inability to modify the QC protocol. This constraint may limit the application of the linearity tests by excluding all Siemens scanners.

V. CONCLUSIONS

Of the quantitative values monitored as part of a QC program, standard deviation of water yielded the highest failure rate and, therefore, is strongly suggested for inclusion in any daily QC program (Table IX). Mean water CT number is easily collected along with standard deviation and is, therefore, also suggested. Measures of uniformity and linearity have relatively low failure rates and, therefore, may not require frequent (daily, weekly, or monthly) monitoring but should be collected as part of annual surveys and after preventative maintenance has been performed. Water phantom artifact analysis is also encouraged as part of a daily QC program due to its high impact on clinical image quality, even in CT scanners equipped with automated daily QC test programs. In summary, mean water CT number, standard deviation of water, and artifact analysis are suggested for inclusion in daily QC programs. This suggestion, based on failure rates calculated in this study as well as the authors' experience, mirrors the daily QC recommendations outlined by the ACR CT accreditation program.¹³

Large phantom artifact analysis had the highest failure rate of all the tests included in this study. While large phantom artifact analysis has a high impact on clinical image quality, the test does require a separate phantom and scan protocol. As such, we recommend this test as part of a weekly or bi-weekly QC program as "best practice" for facilities possessing an appropriate phantom to ensure adequate image quality across the entire scanner field of view.

TABLE IX. Test and frequency suggestions for a technologist CT QC program.

Test	Daily	Weekly/monthly	Annually ^a	After service
Water, mean CT number	X		X	X
Water, standard deviation	X		X	X
Water, uniformity			X	
Linearity, water ^b			X	
Linearity, acrylic ^b			X	
Linearity, air ^b			X	
Water phantom artifact	X		X	X
Large phantom artifact		X	X	X
Retrospective QC analysis			X	

^aShould be performed in concert with a qualified medical physicist.

^bOther standard materials can be used as desired.

While regular QC is invaluable to any clinic, time limitations can make a full QC program impractical at a smaller clinic where the entire process must be done manually by CT technologists. The QC process can be expedited by taking advantage of available automatic tools for quantitative test analysis and the use of stack mode for artifact analysis by an experienced technologist. Because the majority of water phantom artifacts are rings, artifact analysis may be further accelerated by the development of automated tools for ring identification.

Regular review of QC results by an appropriate QC team is encouraged to help identify and resolve recurring scanner issues. Retrospective analysis on an annual scale and even longer is encouraged for all quantitative values included in the daily QC program as it aids in identification of trends and verification of appropriate test limits.

ACKNOWLEDGMENTS

The authors would like to acknowledge the MD Anderson Cancer Center physics and CT technologists for their dedication to routine QC, which allowed for the collection and analysis of the QC data used in this study.

^aAuthor to whom correspondence should be addressed. Electronic mail: dcody@mdanderson.org

¹D. D. Cody, D. M. Stevens, and J. Rong, "CT quality control," in *Advances in Medical Physics – 2008*, edited by A. B. Wolbarst, K. L. Mossman, and W. R. Hendee (Medical Physics, Madison, 2008), pp. 47–60.

²Minnesota Administrative Rules: Computed Tomography Requirements, <https://www.revisor.mn.gov/rules/?id=4732.0860>.

³New Jersey Compliance Guidance for Computed Tomography Quality Control, New Jersey Department of Environmental Protection, Bureau of Radiological Health, <http://www.state.nj.us/dep/rpp/download/ctcgd.pdf>.

⁴New York State Department of Health, Bureau of Environmental Radiation Protection, Guide for Radiation Safety/Quality Assurance Programs: Computed Tomography Equipment, http://www.health.state.ny.us/environmental/radiological/radiation_safety_guides/catguide.htm.

⁵CRCPD, "Suggested State Regulations for the Control of Radiation (SS-RCR), Part F," *Conference of Radiation Control Program Directors*, Frankfurt, KY (2001), <http://www.crcpd.org/SSRCRs/F-Part%202009.pdf>.

⁶AAPM, "Quality Control in Diagnostic Radiology: Report of Task Group No. 12, Diagnostic X-Ray Imaging Committee," American Association of Physicists in Medicine Report No. 74, Madison, WI (Medical Physics, 2002).

- ⁷ACR, American College of Radiology CT Accreditation Program Requirements, <http://www.acr.org/~media/ACR/Documents/Accreditation/CT/Requirements.pdf>.
- ⁸Computed Tomography Radiation Safety Issues in Ontario, developed by Healthcare Human Factors Group, Centre for Global eHealth Innovation, University Health Network, Toronto, ON, Canada. http://www.ehealthinnovation.org/files/CT_radiation_safety.pdf.
- ⁹EUR 16262, European Guidelines on Quality Criteria for Computed Tomography, <http://www.drs.dk/guidelines/ct/quality/index.htm>.
- ¹⁰IPEM, "Recommended Standards for the Routine Performance Testing of Diagnostic X-Ray Imaging Systems," Institute of Physics and Engineering in Medicine Report 91, York, United Kingdom (IPEM, 2005).
- ¹¹C. H. McCollough and F. E. Zink, "Quality control and acceptance testing of CT systems," in *Medical CT and Ultrasound: Current Technology and Applications*, AAPM Summer School Proceedings, edited by L. W. Goldman and J. B. Fowlkes (Advanced Medical, Madison, 1995), pp. 437–465.
- ¹²J. Papp, *Quality Management in the Imaging Sciences*, 3rd ed. (Mosby, St. Louis, 2006).
- ¹³Computed tomography quality control manual, American College of Radiology, Reston, VA, 2012.
- ¹⁴J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems*, 2nd ed. (Prentice-Hall, New York, 2006).
- ¹⁵AAPM, "Quality Assurance Methods and Phantoms for Magnetic Resonance Imaging: Report of Task Group No. 1, Nuclear Magnetic Resonance Committee," American Association of Physicists in Medicine Report No. 28, Madison, WI (Medical Physics, 1990).
- ¹⁶B. W. Murphy, P. L. Carson, J. H. Ellis, Y. T. Zhang, R. J. Hyde, and T. L. Chenevert, "Signal-to-noise measures for magnetic resonance imagers," *Magn. Reson. Imaging* **11**, 425–428 (1993).
- ¹⁷K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika* **73**, 13–22 (1986).
- ¹⁸D. D. Cody, D. M. Stevens, and L. E. Ginsberg, "Multi-detector row CT artifacts that mimic disease," *Radiology* **236**, 756–761 (2005).