Gain and offset calibration reduces variation in exposure-dependent SNR among systems with identical digital flat-panel detectors

Charles E. Willis,^{a)} Yevgeniy Y. Vinogradskiy, and Brad K. Lofton
*Department of Imaging Physics, The University of Texas M. D. Anderson Cancer Center,
Houston, Texas 77030-3721*

R. Allen White
*Department of Bioinformatics and Computational Biology, The University of Texas M. D. Anderson Cancer
Center, Houston, Texas 77030-3721*

(Received 21 October 2010; revised 23 May 2011; accepted for publication 3 June 2011; published 30 June 2011)

Purpose: The conditions under which vendor performance criteria for digital radiography systems are obtained do not adequately simulate the conditions of actual clinical imaging with respect to radiographic technique factors, scatter production, and scatter control. Therefore, the relationship between performance under ideal conditions and performance in clinical practice remains unclear. Using data from a large complement of systems in clinical use, the authors sought to develop a method to establish expected performance criteria for digital flat-panel radiography systems with respect to signal-to-noise ratio (SNR) versus detector exposure under clinical conditions for thoracic imaging.

Methods: The authors made radiographic exposures of a patient-equivalent chest phantom at 125 kVp and 180 cm source-to-image distance. The mAs value was modified to produce exposures above and below the mAs delivered by automatic exposure control. Exposures measured free-in-air were corrected to the imaging plane by the inverse square law, by the attenuation factor of the phantom, and by the Bucky factor of the grid for the phantom, geometry, and kilovolt peak. SNR was evaluated as the ratio of the mean to the standard deviation (SD) of a region of interest automatically selected in the center of each unprocessed image. Data were acquired from 18 systems, 14 of which were tested both before and after gain and offset calibration. SNR as a function of detector exposure was interpolated using a double logarithmic function to stratify the data into groups of 0.2, 0.5, 1.0, 2.0, and 5.0 mR exposure (1.8, 4.5, 9.0, 18, and 45 μ Gy air KERMA) to the detector.

Results: The mean SNR at each exposure interval after calibration exhibited linear dependence on the mean SNR before calibration ($r^2 = 0.9999$). The dependence was greater than unity ($m = 1.101 \pm 0.006$), and the difference from unity was statistically significant ($p < 0.005$). The SD of mean SNR after calibration also exhibited linear dependence on the SD of the mean SNR before calibration ($r^2 = 0.9997$). This dependence was less than unity ($m = 0.822 \pm 0.008$), and the difference from unity was also statistically significant ($p < 0.005$). Systems were separated into two groups: systems with a precalibration SNR higher than the median SNR ($N = 7$), and those with a precalibration SNR lower than the median SNR ($N = 7$). Posthoc analysis was performed to correct for expanded false positive results. After calibration, the authors noted differences in mean SNR within both high and low groups, but these differences were not statistically significant at the 0.05 level. SNR data from four additional systems and one system from those previously tested after replacement of its detector were compared to the 95% confidence intervals (CI) calculated from the postcalibration SNR data. The comparison indicated that four of these five systems were consistent with the CI derived from the previously tested 14 systems after calibration. Two systems from the paired group that remained outside the CI were studied further. One system was remedied with a grid replacement. The nonconformant behavior of the other system was corrected by replacing the image receptor.

Conclusions: Exposure-dependent SNR measurements under conditions simulating thoracic imaging allowed us to develop criteria for digital flat-panel imaging systems from a single manufacturer. These measurements were useful in identifying systems with discrepant performance, including one with a defective grid, one with a defective detector, and one that had not been calibrated for gain and offset. The authors also found that the gain and offset calibration reduces variation in exposure-dependent SNR performance among the systems. © 2011 American Association of Physicists in Medicine. [DOI: 10.1118/1.3602458]

Key words: digital radiography, quality control, SNR, flat-panel detector

I. INTRODUCTION

The absence of performance standards for digital radiography systems presents a practical challenge for clinical medical physicists. The conditions under which vendor performance criteria are obtained often do not adequately simulate actual clinical imaging conditions with respect to radiographic technique factors, scatter production, and scatter control. For example, one vendor's exposure condition for an upright exposure station specifies an 80 kVp exposure through a 2 cm Al filter mounted on the collimator with the grid removed. While this condition may adequately represent the effective energy of the x-ray beam exiting a patient in a chest examination, it does not simulate the scatter-to-primary ratio of the beam incident on the image receptor in a clinical examination, nor does it mimic the presence of a grid. Therefore, it is difficult to assess whether performance under ideal conditions is indicative of performance under clinical conditions. It should be possible to collect data under more realistic conditions of exposure and to determine whether the device is functioning as expected. A wide variety of patient-equivalent phantoms are readily available to simulate the attenuation and scatter properties of human anatomy in projection radiography. A higher kilovolt peak that is more relevant to conventional thoracic radiography can be used with these phantoms.

A reasonable expectation is that identical imaging devices should perform similarly to each other under identical conditions of exposure. For example, under identical clinical exposure conditions simulating attenuation and scatter in the patient, as well as scatter removal, the signal-to-noise ratio (SNR) as a function of detector exposure should be the same for identical digital flat-panel detectors. Wide variation in exposure-dependent SNR among identical systems is likely to cause inconsistent presentation of images for interpretation, especially considering that the benefits of digital image processing may be limited by the SNR within the image.

Toward this end, using data from a large complement of systems in clinical use, we sought to develop a method to establish expected performance criteria for digital flat-panel radiography systems with respect to SNR versus detector exposure under clinical conditions for thoracic imaging. The writing that follows describes this work and our observations of how calibration of gain and offset affects variation among systems.

II. METHODS AND MATERIALS

Figure 1 shows the test setup for imaging of the patient-equivalent phantom (CDRH LucAl Chest)¹ using the upright exposure station. The phantom comprised uniform slabs of materials as described in the reference cited. Any apparent nonuniformity is surface oxidation of the Al layer, which was deemed inconsequential for radiography. Radiographic exposures of the LucAl Chest phantom were made at 125 kVp and 180 cm source-to-image distance (SID) using 18 digital radiographic systems (including models Revolution XQi, XR/d, XR/dII, and Definium; GE Medical Systems; Milwaukee, WI). All of these systems have an indirect digital flat-panel detector with a CsI(Tl) conversion layer and

Fig. 1. Test setup. Patient-equivalent phantom (LucAl Chest) imaged using an upright exposure station with measurement of free-in-air entrance exposure according to AAPM Report No. 31 (Ref. 3). Apparent nonuniformity is surface oxidation of the Al.

$200 \times 200 \mu\text{m}$ detector elements (del).² Nine of the systems had dual detectors; in these systems, only the upright exposure stations were tested. Models XQi, XR/d, and XR/dII had 180 cm focused, 13:1, 78 line/cm, Al interspace grids, with $29 \mu\text{Pb}$ thickness. Definiums had 180 cm focused, 13:1, 70 line/cm, Al interspace grids, with $40 \mu\text{Pb}$ thickness. We modified the mAs value to produce exposures above and below the mAs delivered by automatic exposure control. Exposure was measured using a 15 cc chamber from a Keithley Triad electrometer (Cardinal Health, Cleveland, OH) that was within 1 yr of calibration. The ion chamber was located in the probe holder of the arm on the patient-equivalent phantom, which positions the chamber approximately 24 cm forward of the front surface of the phantom to minimize backscatter. Exposure measured at this point is considered to be "free-in-air."³ Exposure measurements were corrected to the imaging plane according to the inverse square law, multiplied by an empirically determined factor of 0.226 to account for attenuation by the phantom and divided by another empirically determined factor of 2.45 to account for the Bucky factor of the grid for this phantom, exposure geometry, and kilovolt peak.

The SNR was calculated as the ratio of the mean to the standard deviation (SD) of an automatically selected region of interest (ROI) approximately 23 mm in diameter in the center of each unprocessed image. The automatic ROI selection and its default size of 23 mm is a feature of the vendor's acquisition station software. The default size was used in order to assure reproducibility. The size of the ROI at the 180 cm SID corresponds to less than 0.4° degrees from the central ray, suggesting that nonuniformity of exposure from the heel effect should be minimal. The ROI corresponded to approximately 10 000 dels. We assumed that the dels are "stationary" (the system is "ergodic"),⁴ i.e., all dels are identical and an average of 10 000 detectors is the same as the average of one detector taken 10 000 times. This is a reasonable simplifying assumption, especially considering that the

dels are of identical design, and that the gain and offset calibration is specifically intended to compensate for nonidentical behavior.

In collecting data, we used the “raw” image file (vendor nomenclature), or the “for processing” image (DICOM nomenclature), to avoid subsequent rescaling, contrast enhancement, and edge restoration that would modify the detector’s linear characteristic function. Data were acquired from 18 systems, 14 of which were tested both before and after a gain and offset calibration. The gain and offset calibration was a semiautomatic function performed in service mode by the field service engineer using exposures at 60 kVp unfiltered, and at 80 and 120 kVp through a 2.0 cm Al-1100 filter attached to the collimator. For the dedicated chest systems (XQi), the stationary grid is not removable so the calibration is performed with the grid in place. All other systems (XRd, XRd/II, and Definium) have removable stationary grids, and the calibration is performed with the grid removed. The purpose of the calibration is to correct for differences in gain and offset among dels.⁵ The calibration is normally performed in the field during commissioning or upon replacement of a detector. Recalibration is performed only when indicated by the manufacturer’s semiautomated quality assurance measurements. While we may assume that a calibration was performed during installation, records of subsequent calibrations were not maintained, and the systems had been in clinical use for a variable amount of time before our study. Thus, the time elapsed between the most recent calibration and the “before” data is unknown.

Operator verification of vendor performance criteria is accomplished by a semi-automated Quality Assurance Procedure (QAP).⁶ The procedure includes two flat-field exposures at 80 kVp through a 2.0 cm Al filter, from which electronic

noise, correlated noise, bad pixel artifacts, global and local brightness nonuniformity, SNR nonuniformity are determined. The procedure also includes an exposure of a composite phantom, from which five values of modulation transfer function, three values of contrast-to-noise ratio, resolution nonuniformity, and dynamic range linearity and accuracy are determined. QAP testing is performed weekly in our institution on all of these systems. QAP testing did not fail any of the systems on any criterion before or after calibration.

We interpolated the SNR versus detector exposure data using a double logarithmic function in order to stratify the data into groups of 0.2, 0.5, 1.0, 2.0, and 5.0 mR exposure (1.8, 4.5, 9.0, 18, and 45 μ Gy air KERMA) to the detector. In order to determine the uncertainty in our exposure-dependent SNR measurements for an individual system, we analyzed data for one system (Definium 3) for which we acquired four sets of data in the same manner described above for the 14 paired systems. Data were analyzed using SPSS PASW[®] Statistics (version 17.0; IBM Corporation, Somers, NY).

III. RESULTS

Paired SNR data for 14 systems before and after calibration are shown as a function of the logarithm of detector exposure in Fig. 2. The relationship between log of SNR and the log of detector exposure was not always linear both before and after calibration, as would be expected for a quantum-limited system where SNR^2 is proportional to the number of incident photons.⁷ Although the log of the mean SNR for all systems was roughly linear (average $r^2 = 0.9654 \pm 0.0267$) with respect to the log of exposure, individual systems deviated from strictly linear log-log dependence. It is important to note

FIG. 2. Box and whisker charts of exposure-dependent signal-to-noise ratio for 14 paired systems (a) before and (b) after calibration. Dark lines indicate the medians, boxes indicate quartiles, and whiskers show the range of data excluding minor (values that are between 1.5 and three times the interquartile range) and major outliers (values that are more than three times the interquartile range). Labeled points in (b) indicate two minor outliers from unit DC and one from Definium 3.

TABLE I. Symbols and coefficients of log (SNR) versus log (detector exposure).

System	Symbol		Before calibration			After calibration		
	Figure 4	Figure 5	m	b	r^2	m	b	r^2
A1	*		0.3247	1.8953	0.9580	0.3706	1.9132	0.9549
A4	Δ		0.2972	1.8795	0.9401	0.4062	1.9308	0.9658
A6	+ (Red)		0.3956	2.0251	0.9736	0.4753	2.0273	0.9831
C1	$\square$		0.4532	2.0713	0.9674	0.4048	1.9339	0.9672
C2	$\diamond$		0.4361	2.0635	0.9787	0.4358	1.9492	0.9777
C3	$\circ$		0.3451	1.9084	0.9617	0.4869	1.9844	0.9870
E1	•		0.4885	2.0022	0.9954	0.4589	2.0040	0.9929
E2	$\blacktriangle$		0.3699	1.9296	0.9625	0.3803	1.9316	0.9698
DC	$\blacksquare$ (Red)		0.4666	2.0192	0.9789	0.4957	2.1035	0.9897
F12	x		0.5173	2.0408	0.9974	0.3649	1.9893	0.9829
I1	$\blacklozenge$		0.3178	1.8905	0.9560	0.5039	2.0207	0.9932
I10	+ (Blue)		0.3089	1.8945	0.9586	0.3595	1.8853	0.9362
I2		Δ	0.3453	1.8979	0.9770	—	—	—
I3		$\square$	—	—	—	0.4796	1.9911	0.9973
ROC		x	—	—	—	0.4075	1.9472	0.9619
Definium 1	$\blacklozenge$ (Green)		0.3979	1.9313	0.9486	0.3767	1.9974	0.9650
Definium 1 (new detector)		$\circ$	—	—	—	0.3819	1.9751	0.9682
Definium 2		$\diamond$	—	—	—	0.4578	2.0077	0.9801
Definium 3	$\blacksquare$ (Blue)		0.2155	1.8790	0.8673	0.2892	1.9013	0.9225
Average			0.3786	1.9552	0.9614	0.4186	1.9718	0.9720
SD (%)			21.823	3.704	3.171	14.041	2.705	2.037
Average (w/out Definium 3)			0.3903	1.9607	0.9681	0.4263	1.9760	0.9749
SD (%)			18.404	3.667	1.699	11.876	2.627	1.636

SNR, signal-to-noise ratio; SD, standard deviation; m, slope; b, intercept; r^2 , coefficient of determination, SD (%) is the standard deviation expressed as a percentage of the mean.

that although the manufacturer makes no claims about the performance of the systems under the conditions of exposure used here, none of the detector exposures exceeded their stated conditions for saturation, and none were outside levels that would be encountered in clinically relevant portions of images during routine clinical use.

Table I identifies 18 systems that were tested including one whose detector was replaced (Definium 1—new detector) and shows the symbols that are used to represent the systems in two figures that appear later. Table I lists the slopes (m), intercepts (b), and coefficients of determination (r^2) for the least squares fits of the logarithm (base 10) of SNR versus the logarithm (base 10) of exposure data collected. Of the 14 systems for which data both before and after calibration were available, the coefficient of determination increased for eight systems after calibration but decreased for six systems after calibration. These results demonstrate that the calibration had no systematic effect on the degree of linearity of the dependence of the logarithm of SNR on the logarithm of exposure. One system, Definium 3, had the lowest coefficient of determination of all systems tested both before and after calibration. Averages and standard deviations were calculated for all systems and excluding Definium 3. The average of the intercepts is about two, which corresponds to an SNR of about 100 at a detector exposure of approximately 1 mR.

The noise-equivalent quanta (NEQ) of an imaging system as a function of exposure and spatial frequency is defined as the square of the SNR in the image.⁷ The frequency-depend-

ent NEQ is related to the product of the frequency-dependent detective quantum efficiency and the square of the SNR of the incident x-ray beam. There are some limitations in applying this definition to flat-panel detectors, with respect to under-sampling.⁸ Our large-area averaged measurement of SNR relates to NEQ at zero spatial frequency. For stimulative phosphor systems, NEQ increases proportionally with the exposure up to 1 mR.⁷ We can expect that NEQ of indirect DR systems should have similar behavior. At low exposure levels, including those used in medical radiography, NEQ is proportional to the number of photons; therefore, the logarithm of SNR (and its square) should be linear with the logarithm of exposure.

For each system tested, a linear least squares fit showed that the mean ROI value was highly linear ($r^2 > 0.99$) with respect to the calculated detector exposure and, coincidentally, was roughly equivalent to the detector exposure in units of μR . For each system tested, a power-law fit indicated that the SD of the ROI increased according to a fractional exponent of the detector exposure that exceeded 0.5 in all but one (0.4885 before calibration) of 36 measurements (including 33 shown in Table I plus three repeated measurements described later). We interpreted this to be an indication that noise sources other than quantum mottle are significant for these systems under these conditions of exposure, with possible sources being scattered radiation and fixed pattern noise. Furthermore, the magnitude of the SD of the ROI was only about 1/3 of the square root of the corresponding mean value.

TABLE II. Mean, standard deviation, and confidence intervals of SNR.

Detector exposure mR (μ Gy)	Before calibration				After calibration			
	Mean SNR	SD	−95%CI	+95%CI	Mean SNR	SD	−95%CI	+95%CI
0.2 (1.8)	49.49	4.42	46.94	52.04	47.98	4.47	45.40	50.56
0.5 (4.5)	70.42	8.86	65.30	75.53	70.30	7.91	65.73	74.87
1.0 (9.0)	92.29	15.90	83.11	101.46	94.04	13.42	86.29	101.79
2.0 (18)	121.33	27.11	105.68	136.99	126.02	22.55	113.00	139.05
5.0 (45)	175.04	51.35	145.40	204.69	186.08	42.96	161.27	210.89

SNR, signal-to-noise ratio; SD, standard deviation; CI, confidence interval.

Further analysis of the exposure-dependent variance of our data might provide additional insight into the identities and relative magnitudes of contributing noise sources. However, the purpose of this work was to establish clinical performance criteria for the exposure-dependent SNR for these systems and to reduce variation among them, without regard to the individual sources of noise. Since all of these systems use the same detector type, the same grids, and are from the same manufacturer, we expected the sources of noise to be similar. Detailed noise analysis was not necessary in order to accomplish our objectives and was not attempted.

The mean and standard deviation of the pre and postcalibration SNR for the 14 paired systems at each detector exposure level are listed in Table II. A paired *t*-test of the mean SNR before and after calibration did not demonstrate a statistically significant difference ($p = 0.227$). As shown in Fig. 3, there was a strong correlation ($r^2 = 0.9999$) between the mean SNR before and after calibration, as well as between the SD of the mean SNR before and after calibration ($r^2 = 0.9997$). The dependence of the mean SNR after calibration on the mean SNR before calibration was greater than unity as indicated by the slope in Fig. 3(a) ($m = 1.101 \pm 0.006$), and the difference from unity was statistically significant ($p < 0.005$). The dependence of the SD of mean SNR after calibration on the SD of the mean SNR before calibration was less than unity, as indicated by the slope in Fig. 3(b) ($m = 0.822 \pm 0.008$), and the differ-

ence from unity was statistically significant ($p < 0.005$). This suggests that, on average, the gain and offset calibration improved SNR and reduced the variation in SNR among these 14 systems. The magnitude of these effects depends on exposure.

The uncertainty of exposure-dependent SNR measurements was determined from repeated measurements of an individual system. The SD as a percent of the mean SNR was slightly dependent on detector exposure ranging from 3% to 6% and had an average value of 5%. This is 2–4 times smaller than the SD of the mean SNR from the group of 14 paired systems before and after calibration. This comparison suggests that, as expected, the variation in SNR measurements for an individual system is less than the variation of a group of systems. The size of symbols in Figs. 4–7 approximate the uncertainty in measurement for an individual system at the mid-exposure level, that is, $\pm 5\%$ of an SNR value of 100, corresponding to a total of 100 in the value of SNR².

Confidence intervals (CI) are reported in Table II and are shown in Figs. 4–7. The CI was calculated by *spss* from interpolation of the measured SNR data as described, which assumes Gaussian variation. If the variation in the data was indeed Gaussian, then certainly fewer systems would be expected to fall outside the 95% CI. Our results, therefore, suggest that the variation among systems is not Gaussian.

FIG. 3. (a) Mean (squares) and (b) standard deviation (SD; triangles) of SNR of 14 paired systems after gain and offset calibration versus before calibration.

FIG. 4. Exposure-dependent SNR data from 14 paired systems (a) before and (b) after calibration compared to 95% CIs (broken lines). Data from individual systems are depicted with the same symbols in both (a) and (b). Systems with data higher than the CI are shown in red, within the CIs are green, and below the CI are blue. Note that the number of systems outside the CIs decreased from 11 to 6 after calibration. The uncertainty of the individual measurements of SNR is $\pm 5\%$ at the 1.0 mR exposure level, which is approximated by the dimensions of the symbols. Symbols correspond to individual systems as defined in Table I.

Even though the CI was narrower than before calibration [Table II, also note relative size of shaded quartiles in Figs. 2(a) and 2(b)], Fig. 4 shows that the data were more conformant with the CI after calibration. The number of systems outside the CI decreased from 11 of 14 to 6 of 14 after calibration. After calibration, exposure-dependent SNR increased in some systems and decreased in others. This suggests that the benefit of gain and offset calibration is not necessarily an improvement in exposure-dependent SNR but rather an equalization of exposure-dependent SNR among individual systems. Equalizing the exposure-dependent SNR among individual systems may be important for producing consistent images for viewing by physicians.

In an effort to test the idea that the gain and offset calibration has the effect of equalizing the exposure-dependent SNR, we separated the systems into two groups: those with an SNR higher than the median SNR ($N=7$) and those with an SNR lower than the median SNR ($N=7$) for at least four of the five exposure levels tested. We compared the two groups at

each exposure level before and after calibration using post-hoc analysis of variance (ANOVA) tests. The post-hoc analysis was performed to correct for expanded false positive results. The results of the Game–Howell method are shown in Table III. Before calibration, high and low groups were significantly different ($p < 0.05$) from each other (case I), except at the lowest exposure level. Calibration did not produce significant differences within the high group (case II). Calibration produced differences within the low group (case III), but these differences were not significant (i.e., $p > 0.05$). After calibration, high and low groups were no longer significantly different from each other (case IV) at the 0.05 level. These analyses show that, except for case I, the differences in the multiple populations were not statistically significant.

In an attempt to improve the statistical power of the test, one extreme system was excluded from each of the two groups: the system labeled “DC” in Table I was removed from the high group and the system labeled “Definium 3”

FIG. 5. Exposure-dependent SNR data for five additional systems compared to the CIs (broken lines) of 14 paired systems after calibration [from Fig. 4(b)]. Note that four of the systems conformed to the postcalibration CIs developed from the other 14 systems. The nonconformant system (I2) was replaced before it could be recalibrated. Symbols correspond to individual systems as defined in Table I.

FIG. 6. Exposure-dependent SNR investigation of low-performing, nonconformant system. Unit Definium 3 is shown before (blue $\square$) and after (blue $+$) calibration. In both cases, the most of the SNR measurements were below the CIs (broken lines) for 14 paired systems postcalibration [from Fig. 4(b)]. With a substitute grid (green $\blacksquare$), Definium 3 conforms to the CIs. Using Definium 3's grid (blue Δ), another unit (Definium 1) fails to conform to the CIs, although it conforms when using its own grid (green $\blacktriangle$).

Fig. 7. Exposure-dependent SNR investigation of high-performing, nonconformant system. Unit DC was consistently higher than the calibrated CIs (broken lines) for the 14 paired systems [from Fig 4(b)]; before ($\square$), after (+ and $\circ$), in high (*), standard ($\diamond$), and low (X) vertical positions, and after grid replacement (Δ). The same system with a new detector (green $\bullet$) conforms to the CIs.

was removed from the low group. The post-hoc ANOVA was repeated and the results are shown in Table III in the column labeled “ $N=12$ ”. High and low groups remained significantly different before calibration except at the lowest exposure level (case I). The differences before and after calibration increased somewhat for the high group (case II) without achieving statistical significance. Within the low group, calibration resulted in differences approaching significance at the lowest exposure level, while other exposure levels were still not significantly different (case III). After calibration (case IV), high and low groups were not significantly different from each other at any exposure levels for $N=12$. These analyses show that, except for case I, the dif-

TABLE 3. Comparison of high and low systems before and after calibration (Games–Howell Method).

Case	Description	Exposure (mR)	<i>p</i> value	
			$N=14$	$N=12$
I	Low versus high before calibration	0.2	0.489	0.328
		0.5	0.011	0.003
		1.0	0.002	0.000
		2.0	0.001	0.000
		5.0	0.000	0.000
II	High before versus high after	0.2	0.990	0.869
		0.5	0.930	0.684
		1.0	0.881	0.645
		2.0	0.855	0.653
		5.0	0.845	0.689
III	Low before versus low after	0.2	0.376	0.058
		0.5	0.584	0.591
		1.0	0.241	0.271
		2.0	0.186	0.195
		5.0	0.169	0.165
IV	Low versus high after calibration	0.2	0.189	0.208
		0.5	0.133	0.270
		1.0	0.195	0.366
		2.0	0.291	0.469
		5.0	0.416	0.582

ferences in the multiple populations were not statistically significantly different.

Exposure-dependent SNR data from the four unpaired systems and data after a detector replacement were compared with the CI shown in Table II. As shown in Fig. 5, four of these five systems conformed to the 95% CI of paired data after calibration. The nonconformant system was replaced with a conventional tomographic system shortly after installation, before clinical use and without calibration. One conformant system was replaced with a newer digital radiographic system shortly after collecting data, without recalibration and further testing. To examine whether these CI are useful for indicating systems that are not functioning as expected, two of the six systems that remained nonconformant after calibration from the 14 paired systems were studied further.

As mentioned previously, definium 3 had the lowest coefficient of determination and the lowest exposure-dependent SNR performance of all systems tested both before and after calibration (see Table I). As shown in Fig. 6, it improved only slightly after calibration. When the grid from “Definium 1” was substituted for the one delivered with Definium 3, the exposure-dependent SNR conformed to the postcalibration CI. Conversely, when Definium 3’s grid was used with Definium 1, the exposure-dependent SNR fell below the CI. Definium 3’s original grid has been replaced by the manufacturer for continued clinical use.

DC had the highest exposure-dependent SNR of all systems tested after calibration. We repeated the measurement, recalibrated the receptor, measured again, and repeated this measurement with a larger field of view (FOV), with each successive result farther above the CI. To examine the possibility of a grid-central axis misalignment, we repeated the measurement at high, standard, and low vertical positions with increasingly abnormal results, respectively. We replaced the grid, which for this specific type of system involves recalibration of the gain compensation map, and repeated our measurements, resulting in an even higher exposure-dependent SNR. This system is one of the oldest in our inventory, dating back to 2001. Automated quality assurance phantom results indicated a decline in the spatial resolution performance nearing lower limits. At the manufacturer’s suggestion, we ordered a new detector. After calibration, the performance of the new detector was well within our CI as shown in Fig. 7.

IV. DISCUSSION

There are several limitations to this study. Our use of the term “identical” in describing these detectors is an exaggeration. Although the technology of the detectors is identical, the age of the detectors span many years, and four of the systems included are the vendor’s newest version, which is reputed to have improved detection efficiency. The x-ray generators for six of the systems in the study are different than the other 12 systems. The individual x-ray generators vary slightly with respect to output, kilovolt peak, and HVL. There is also some variation in the specific versions of devices and software within the systems that control output. Our results indicate

that these variations are minor compared to the overall effects of gain and offset calibration on exposure-dependent SNR for these systems. It is likely that generator performance characteristics are more readily standardized and controlled than are the gain maps of the associated detector.

The comparisons in Table III require further discussion. Case I confirms what we would expect: grouping systems above and below the median results in samples that are significantly different. Cases II and III confirm the previous findings that calibration does not cause statistically significant changes in the mean SNR at any exposure level. This is likely because the change in the mean SNR is small compared to the large SD of the distributions: recall that the distributions appear to be broader than Gaussian. Case IV demonstrates that after calibration, the two groups are no longer significantly different. This might have been accomplished by random variation in system performance during the interval between the two tests. This is highly unlikely considering the small uncertainty (3%–6%; average 5%) indicated by repeated measurements of an individual system.

We have no data to determine whether the gain and offset calibrations degraded over time or whether they were performed improperly or not at all during initial installation of the systems. Further study is needed to obtain longitudinal results of exposure-dependent SNR for these systems.

We also investigated whether the CI determined under conditions simulating routine examinations of the thorax would have any relevance to exposure-dependent SNR for other examinations. As stated in the methods, all the data used to determine CI were collected from upright exposure stations. Nine of the systems tested also included a second identical detector in a table configuration. We collected additional data from four of these table detectors using two modified protocols, i.e., 100 cm SID and 125 kVp using the LucAl Chest phantom, and 100 cm SID and 80 kVp using another patient-equivalent phantom (the LucAl Abdomen).⁹ Each set of data was analyzed as before to determine detector exposure using appropriate empirical values for attenuation and Bucky factor for the 100 cm SID, and for the particular phantom. These exposure-dependent SNR data fell consistently below the lower CI established for the calibrated upright exposure stations, so we plan to recalibrate our table detectors. The result of these recalibrations is the topic for a later report.

Finally, because all data were obtained from systems from a single manufacturer, our observations have limited applicability to commercial systems from other manufacturers with other detectors. It is reasonable to expect that a similar method could be applied to other manufacturers' systems, and it is reasonable to expect that the specific numerical values of exposure-dependent SNR and CI would be

different. We plan to extend this work to devices from other vendors.

V. CONCLUSIONS

A method was developed for determining the exposure-dependent SNR of a digital flat-panel radiography system under conditions simulating those in thoracic imaging. The exposure-dependent SNRs differed among identical digital flat-panel radiographic systems from a single manufacturer. We found that gain and offset calibration reduces variation among systems. Confidence intervals established for a collection of these systems by this method provide objective criteria for evaluating systems on the basis of their exposure-dependent SNR. When the exposure-dependent SNR of a system is determined to fall outside the 95% CI, we recommend that a gain and offset calibration should be performed. If calibration is unsuccessful in restoring performance to values within the established CI, further investigation may reveal underlying causes of discrepant performance, including the grid and detector.

^{a)}Author to whom correspondence should be addressed. Electronic mail: chwillis@mdanderson.org; Telephone: (713) 563-2721; Fax: (713) 792-6495.

¹B. J. Conway, P. F. Butler, J. E. Duff, T. R. Fewell, R. E. Gross, R. J. Jennings, G. H. Koustenis, J. L. McCrohan, F. G. Rueter, and C. K. Showalter, "Beam quality independent attenuation phantom for estimating patient exposure from x-ray automatic exposure controlled chest examinations," *Med. Phys.* **11**, 827–832 (1984).

²H. G. Chotas, J. T. Dobbins III, and C. E. Ravin, "Principles of digital radiography with large area, electronically readable detectors: a review of the basics," *Radiology* **210**, 595–599 (1999).

³"Standardized methods for measuring diagnostic x-ray exposures," AAPM Report No. 31, July 1990, 22 pages.

⁴J. H. Siewerdsen, L. E. Antonuk, Y. El-Mohri, J. Yorkston, J. M. Boudry, and I. A. Cunningham, "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology," *Med. Phys.* **24**, 71–89 (1997).

⁵J. A. Seibert, J. M. Boone, and K. K. Lindfors, "Flat-field correction technique for digital detectors," *Proc. SPIE*, **3336**, 348–354 (1998).

⁶B. Belanger, J. M. Boudry, A. Katakam, and V. S. Polkus, "Advances in and specifications for fluoroscopic imaging systems," in *Specifications, Performance Evaluations, and Quality Assurance of Radiographic and Fluoroscopic Systems in the Digital Era*, edited by L. W. Goldman and M. V. Yester, American Association of Physicists in Medicine Medical Physics Monograph No. 30 (Medical Physics, WI, Madison, 2004), pp. 69–102.

⁷W. Hillen, U. Schiebel, and T. Zaengel, "Imaging performance of a digital storage phosphor system," *Med. Phys.* **14**, 744–751 (1987).

⁸J. T. Dobbins III, "Effects of undersampling on the proper interpretation of modulation transfer function, noise power spectra, and noise equivalent quanta of digital imaging systems," *Med. Phys.* **22**, 171–181 (1995).

⁹B. J. Conway, J. E. Duff, T. R. Fewell, R. J. Jennings, L. N. Rothenberg, and R. C. Fleischman, "A patient-equivalent attenuation phantom for estimating patient exposures from automatic exposure controlled x-ray examinations of the abdomen and lumbo-sacral spine," *Med. Phys.* **17**, 448–453 (1990).