Inter- and intra-scanner variations in four magnetic resonance imaging image quality parameters

Peltonen, Juha I.

2020-12-04

Peltonen , J I , Mäkelä , T , Lehmonen , L , Sofiev , A & Salli , E 2020 , ' Inter- and intra-scanner variations in four magnetic resonance imaging image quality parameters ' , Journal of Medical Imaging , vol. 7 , no. 6 , 065501 . https://doi.org/10.1117/1.JMI.7.6.065501

http://hdl.handle.net/10138/326067
https://doi.org/10.1117/1.JMI.7.6.065501

unspecified
acceptedVersion

Inter- and intra-scanner variations in four MRI image quality parameters

Juha I. Peltonen, D.Sc. (Tech.) [a,*], Teemu Mäkelä, Lic.Phil. [a,b], Lauri Lehmonen, M.Sc. [a,b], Alexey Sofiev, B.Sc. [a,b], Eero Salli, D.Sc. (Tech.) [a]

[a] HUS Medical Imaging Center, Radiology, University of Helsinki, Helsinki University Hospital, P.O. Box 340, FI-00029 HUS Helsinki, Finland

[b] Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland

Abstract.

Purpose: In addition to less frequent and more comprehensive tests, quality assurance (QA) protocol for a magnetic resonance imaging (MRI) scanner may include cursory daily or weekly phantom checks to verify equipment constancy. With an automatic image analysis workflow, the daily QA images can be further used to study scanner baseline performance and both long and short-term variations in image quality. With known baselines and variation profiles, automatic error detection can be employed.

Approach: Four image quality parameters were followed for 17 MRI scanners over six months: **s**ignal-to-noise ratio, image intensity uniformity, **g**hosting artefact and geometrical distortions. Baselines and normal variations were determined. An automatic detection of abnormal QA images was compared with image deviations visually detected by human observers.

Results: There were significant inter-scanner differences in the QA parameters. In some cases, the results exceeded commonly accepted tolerances. Scanner field strengths, or a unit being stationary versus mobile, did not have a clear relationship with the QA results.

Conclusions: The variations and the baseline levels of image QA parameters can differ significantly between MRI scanners. Scanner specific error thresholds based on parameter means and standard deviations are a viable option to detect abnormal QA images.

Keywords: Magnetic Resonance Imaging, quality assurance/control (QA/QC), automatic workflow, medical image analysis

***Corresponding author e-mail: juha.peltonen@hus.fi

1 Introduction

A quality assurance (QA) program of a magnetic resonance imaging (MRI) scanner typically includes a less frequent comprehensive testing and a more frequent, for example daily or weekly, verification of image quality stability [1]. For comprehensive testing, well established and generally accepted methods and error levels exist [2-6]. Daily QA methods are usually faster and more straightforward to integrate with clinical routine, but less informative and anomalies may be harder to interpret.

Imaging a simple homogeneous phantom is a common practice in daily QA. The aim is to verify the proper working order of an MRI scanner prior to the first patient study [7]. Keeping parameters and phantom positioning constant allows an automatic analysis workflow to be used in monitoring the long- and short-term stability of the scanner performance [8]. Ideally, this would allow the QA specialist to determine performance baselines and tolerances, and subsequently detect equipment faults before they significantly impact image quality.

It is important to distinguish an abnormal event from the normal variations of the measured parameters. Inherent variations can be introduced by alterations in phantom placement, fluid movements or temporal changes in the phantom contents, environmental factors (e.g. humidity), or hardware fluctuations. Normal variations in MRI QA parameters have been reported in previous publications. These studies have concentrated mostly on comprehensive testing based on multiple MRI sequences and standardized phantoms [9-12] and on the variability and the long-term behavior of image QA parameters [8, 13-16]. Technical methodologies for QA workflows have been presented by multiple authors [7, 8, 17-19].

In this study, we investigated the variations and normal baselines of four image quality parameters. The measurements were based on daily single image phantom acquisitions. The parameters were followed for 17 MRI systems, including scanners with 1.5 T and 3 T field strengths in stationary and mobile installations. Additionally, the automatic analysis results were compared against visual estimations to study the possibility of incorporating error detection in the workflow results.

2 Material and Methods

2.1 Analysis pipeline

Table 1. The daily QA sequence parameters.

Parameter	Value
Sequence type	Spin echo
TE	20 ms
TR	500 ms
FOV	250 mm x 250 mm
Matrix	256 x 256
Flip angle	90°
Slice thickness	5 mm
Slices	1
Phase encoding direction	R-L
Bandwidth	1.5T: 70 Hz/px 3.0T: 100 Hz/px
Parallel imaging	off
Image filters	off
Image normalization [1)]	on

[1)] *Based on element sensitivity in multi-channel coils*

On all the studied MRI systems, a QA phantom was scanned first in the morning if patient studies were scheduled for the day. A single transversal image slice from the homogenous part of the cylindrical or spherical phantom provided by the scanner manufacturer was acquired. The phantom position was fixed by a compatible phantom holder inside a head coil. Scanning was carried out by using a spin echo sequence with parameters presented in Table 1. The primary purpose of acquiring the phantom image was to verify that the scanner is operational before patient examinations. After visual inspections, the images were sent to a QA server for detailed analysis. The analysis pipeline calculated signal-to-noise ratio (SNR), image intensity uniformity, ghosting artefact magnitude and phantom dimensions. Finally, the results were presented on a hospital intranet web page. Examples of typical time series are shown in Fig. 1.

The SNR was calculated according to the preferred method for a single image in NEMA-MS-1 SNR standard [5]:

$$SNR = 0.66 \times \frac{signal}{noise}. \quad (1)$$

The *signal* was defined as the mean intensity in a circular region of interest (ROI) centered with 80% radius of the phantom's signal producing area. The *noise* was determined by calculating the intensity standard deviation (SD) in the combined area of rectangular background ROIs (Fig. 2a). The factor 0.66 compensates the theoretical Rician distribution of a magnitude image to correspond to an underlying Gaussian distribution [6].

Image intensity uniformities were calculated with three methods presented in NEMA-MS-3 guidance for image uniformity measurements [6] and in IEC standard 62464-1 [4]. The uniformities were determined from the same signal area as in the SNR calculation.

In the method presented by National Electrical Manufacturers Association (NEMA), the uniformity is calculated by

$$Uniformity_{NEMA} = 1 - \frac{S_{\max} - S_{\min}}{S_{\max} + S_{\min}}, \quad (2)$$

where S_{max} and S_{min} are the maximum and minimum intensities in the signal ROI (Fig. 2b). Alternatively, the image may be filtered with the Gaussian kernel

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (3)$$

to minimize the effect of noise. The filtered version is referred hereafter to as NEMA filtered uniformity.

According to the IEC standard 62464-1, the image uniformity is calculated by

$$Uniformity_{IEC} = 1 - \frac{\sum_{1=1}^{N} (|S_i - S|) / N}{S}, \quad (4)$$

where S_i is an individual pixel value inside the signal ROI, S is the mean value of all pixels in the signal ROI, and N is the total number of pixels in the signal ROI.

The image ghosting measurement followed the IEC standard 62464-1 [4]. The signal ROI placement was identical to the SNR calculation. The ghosting ROIs were placed outside the signal producing area in the phase encoding direction (Fig. 2c). The image was first filtered with a 5×5 averaging kernel after which the ghosting percentage was calculated by

$$Ghosting = 100\% \cdot \frac{I_G}{S_G}, \quad (5)$$

where I_G is the highest intensity within the ghosting ROIs and S_G is the mean signal intensity in the signal producing area.

The geometrical distortions were calculated by measuring the longest dimensions of the signal producing area in the vertical and horizontal directions (Fig. 2d). Before the calculation, the phantom image was binarized using a threshold of the mean intensity in the phantom signal producing area divided by two.

2.2 Scanner comparison

The QA parameters were followed for 17 MRI scanners from three vendors. Four of the scanners were stationary 3T scanners (IDs 1–4), three were mobile 1.5T scanners (IDs 5–7), nine were stationary 1.5T scanners (IDs 8–16) and one was a stationary peripheral 1.5T scanner (ID 17).

The scanner performances were compared by calculating mean, median, SD and the coefficient of variation (CV) of each parameter in the time series. The CV was calculated by

$$CV = \frac{SD}{mean}. \quad (6)$$

The absolute SNR values were not comparable because of differences in the scanner models, coils, phantoms, hardware components and installations. Thus, the CVs were used to compare the SNRs between the scanners. Similarly, the phantoms had varying diameters and therefore the CVs were used for the inter-scanner comparison of the geometric distortions.

The scanner differences were assessed with the modified Z-parameter [20]. The comparison was done between all the CVs and between the median values of the image intensity uniformity and ghosting. A modified Z-score above 3.5 or below -3.5 was considered significantly differing.

No major faults were identified during the six months period that could have considerably affected the measured parameters. Before further investigation, any outliers resulting from phantom misalignment or similar gross user-related errors were excluded. The outlier filtering was done by rejecting samples more than four SDs from the mean of each QA parameter. If an image was determined as an outlier with respect to one parameter, it was completely removed from the analysis. Thus, 42 (or 2%) of 2106 images were removed. Twenty-five of these had increased noise or decreased uniformity due to poor coil connections, nine showed phantom misalignment, three included excessive fluid movements, three had artefacts related to air bubbles inside the phantoms and two images were scanned with wrong imaging parameters.

2.3 Scanner stability

QA parameter short-term stability was studied by repeating the daily QA image acquisition 50 times on two scanners (ID 4 and ID 16). The 50 repeats were performed consecutively in a single session without phantom repositioning or shimming. The scanners came from the same manufacturer and were installed at the same time in nearby rooms. Possible transient environmental factors could therefore affect both scanners in a similar fashion. The main field strength of the scanners were 3 T and 1.5 T, respectively. The resulting image quality parameter means and SDs were compared with the values obtained from the six-month test period.

2.4 Abnormal image detection

The feasibility of automatically detecting abnormal images was studied by comparing the automatic QA results with those from human observers. All images were labeled abnormal or normal with respect to SNR, image intensity uniformity, ghosting or geometric distortion by two

experienced QA specialists. The specialists were medical physicists JIP and TM with respective nine and seven years of experience in MRI QA. Based on subjective evaluation, an image was labeled abnormal if it differed from typical images from the specific scanner. The final labeling was the conjunction of both observers' labeling. The inter-observer agreement was studied by calculating Cohen's kappa between the annotations [21].

The labeling was done with an in-house MATLAB (MathWorks, Natick, MA, USA) application. In the user interface, the QA image was presented with three windowing settings: "default", "narrow" highlighting background noise, and "wide" maximizing the dynamic range in the signal producing area. Additionally, the rater could freely alter the window setting.

Finally, the receiver operating characteristic (ROC) curves were calculated between the automatic and visual assessments over the whole data by varying the automatic decision threshold. In the ROC calculation, the threshold was defined as the difference (in SDs) from the QA parameter mean value. Two-sided detection thresholds were used for the SNR and geometric distortion and one sided for the image ghosting and intensity uniformity. Finally, the areas under the ROC curves (AUCs) were calculated.

2.5 User interface

The QA results for each MRI system were communicated to all user groups by the user interface presented in Fig. 3. The interface includes an interactive DICOM viewer and six scatter plots showing the results over the past year. The scatter plots include SNR, ghosting, uniformity (IEC and NEMA) and geometric distortion time series. The DICOM viewer uses open source Cornerstone Core JavaScript library allowing panning, zooming and windowing [22]. Dcmdump tool of the DCMTK library [23] was used for reading the DICOM header to identify the site. The interactive plots use open source JavaScript charting library dygraphs [24]. Each point in the plot represents a single day and when selected, the corresponding image and information are shown. Additionally, the user has a possibility to download the DICOM image for further analysis.

3 Results

The SNR CVs were relatively uniform across the scanners apart from IDs 2 and 4, which had significantly differing modified Z-scores. The CVs are presented in Fig. 4. The average mean SNR CV was 4.9%. According to the ROC curves in Fig. 5, the automatic detection of abnormal SNR had 0.86 AUC. Two-sided error bounds of ±1.3 SDs from the mean value would produce 71% sensitivity with 90% specificity.

The image intensity uniformity medians, ranges and interquartile ranges for the three different methods are presented in Fig. 6. All the mean uniformities were between 94–99% for the IEC method and 82–98% for the NEMA methods. Scanner IDs 2 and 4 differed significantly from the rest using the IEC method and IDs 3 and 4 differed significantly using both NEMA and NEMA filtered methods. In general, the IEC method produced numerically higher uniformity values (98.2% mean) compared to the NEMA and NEMA filtered methods (93.0% and 94.2% means, respectively). Also, the average SD of the IEC method was lower (0.27 %) compared to the NEMA methods (1.50% and 0.90%). According to the ROC in Fig. 5, the optimal abnormal image intensity uniformity detection was achieved with the NEMA method (AUC 0.92). AUCs for NEMA filtered and IEC methods were 0.91 and 0.89, respectively. The NEMA method was producing approximately 87% sensitivity with 90% specificity when using a single sided low error bound of 0.8 SDs from the mean value.

The ghosting median, range and interquartile range are presented in Fig. 7. The means were 0.6–2.9% with an average mean of 1.2%. SDs varied between 0.06–0.8% with an average SD of 0.23%. Scanner ID 3 had significantly differing modified Z-scores for both the ghosting median and CV. Scanner ID 11 had a significantly differing modified Z-score only for CV. According to the ROC analysis (Fig. 5), the ghosting detection AUC was 0.71. The use of a one sided high error bound of 0.5 SDs from the mean value would produce approximately 53% sensitivity with 80% specificity.

The geometric distortion CVs were 0.2–2.5% in the horizontal and 0.2–2.7% in the vertical direction, with the average mean values of 0.5% and 0.7% respectively (see Fig. 8). Scanners ID 2 and ID 9 had significantly differing modified Z-scores in both horizontal and vertical measurements. The QA specialists did not detect geometric distortions visually and ROC analysis was not performed.

The image quality mean values were similar between the 50 consecutive scans in one session and the six-month test period. However, the SDs of the QA parameters from the six-month period were higher. A comparison for the SNR and ghosting is presented in Fig. 9. A full comparison is presented in Appendix Table 2.

The number of images labeled abnormal by both human observers due to SNR, image intensity uniformity and ghosting were 0.7%, 2.5%, 4.9%, respectively. The percentages of images labeled abnormal are presented in Table 2. The respective Cohen's kappas between the observers were 0.97, 0.93, and 0.75.

Table 2. Percentages of images labeled abnormal by the human observers. Geometric distortions were not included as the observers did not detect any faults.

	Observer 1	Observer 2	Observer 1 and Observer 2
SNR	0.9 %	2 %	0.7 %
Image intensity uniformity	4.9 %	3.8 %	2.5 %
Ghosting	6.3 %	16.2 %	4.9 %

The full results considering the scanner specific means, medians, SDs and CVs with respect to all image quality parameters are available in Appendix Tables 3-5. Additionally, a comparison between 3T, 1.5T and mobile 1.5T scanners is available in Appendix Table 6.

4 Discussion

In this study daily QA phantom images from 17 MRI systems were collected for six months. The imaging parameters were standardized. In the presented results, it is evident that individual scanners may produce substantially different results in one or more QA parameters than the rest. The scanners producing divergent results did not distinctly belong to any group based on scanner field strength or mobility, neither there were any technical malfunctions identified. It appears that individual installations are inherently unique with respect to QA parameter baselines and variations.

The absolute SNR values were not comparable between the systems due to different hardware (e.g. coil type) and software solutions. In addition to measured SNR values, this can be seen in greatly varying background noise textures depending on the scanner manufacturer and model (Fig. 10). The differences in noise appearances may have a significant impact on the results and decrease the comparability of the CVs. This is an additional motivation for establishing scanner-specific normal variation levels. Increased comparability of the results could be achieved if the raw data was available and the reconstruction could be calculated identically independent of the vendor. However, most users do not have this option and it could further complicate the workflow.

The image intensity uniformity was measured using three methods. The IEC method is based on the average of the absolute deviation from the mean, whereas the NEMA methods incorporate only the highest and the lowest pixel values. Based on the means and SDs, there is a bias between the absolute values and sensitivity of the tests. Generally, the NEMA methods produced lower absolute image uniformity values with higher SDs. Also, they have greater ROC AUC than the IEC method indicating a better correspondence to the visual inspections. It is likely, that the methods are sensitive to different image artefacts and noise textures. A large interquartile range in the NEMA uniformity test does not necessarily mean a large interquartile range in the IEC uniformity.

Scanners ID 3 and ID 11 had significantly higher CVs in the ghosting measurements compared to the other scanners. ID 3 also had a significantly higher ghosting level, which exceeded the 2.5% acceptance threshold presented in the ACR accreditation guidance [2]. The reason for the increased ghosting in these installations is not known.

The CVs of the geometric distortions in scanners ID 2 and ID 9 stood out from the group. ID 9 had a spherical phantom which was more sensitive to the phantom positioning compared to a cylindrical phantom. Scanner ID 2, however, had a cylindrical phantom. One explanation could be that the vendors' standard QA phantoms may not remain completely rigid between the daily scans as they were not meant for geometric distortion measurements. Thus, a small increase in a CV may not result from a degraded performance or stability. Substantially fluctuating values may, however, indicate that a scanner should not be used for patient studies requiring high geometric accuracy.

According to the ROC curves, the measured parameters had from excellent to fair AUC when compared with consensus labeling by the QA specialists. The differences can be explained by the

subjectivity and difficulty of visually detecting deviating images. According to the AUC, the NEMA uniformity measurement had the best agreement with QA specialists, followed by the rest of the uniformity measures, SNR and ghosting. The good performance in the SNR error detection was unexpected since the visual detection of a decreased SNR is highly subjective.

The high AUC values for all the measured parameters enable the use of an automatic detection of abnormal QA images. The detection thresholds can be set individually based on the mean and SD of each parameter on each scanner. A more detailed setup has been presented by Simmons et al. [17] with additional rules on consecutive measurements to further improve error detection. It is also important to note that the detection of a statistically abnormal image does not automatically mean a hardware problem. It would require additional study to find correlation between individual hardware failures and symptomatic effects in image quality parameters. Thus, in addition to statistical detection thresholds, it is useful to follow presented acceptance thresholds, such as 2.5% ghosting error by ACR [2]. When adopting error bounds from QA standards or publications, the compatibility of the used MRI sequence needs to be taken into consideration. The effect of MRI sequence type on QA parameters has been demonstrated by Peltonen et al. [8]

The correlation between QA specialists in image artefact labeling was from good to excellent. Although a human observer can detect anomalies in QA images, it is hard to define subjective thresholds for any artefact type. Thus, a repeatable and objective automatic abnormal QA image detection system is a valuable tool for a systematic MRI QA.

Characteristic hardware stability in the daily QA measurement was studied by repeating the scan consecutively 50 times with two scanners. The mean values of the image quality parameters were similar to the six-month test on both scanners and, as expected, the SDs were higher in the six-month test. The SDs obtained with 50 consecutive scans can be depicted as optimal variations since they include only minimal short-term scanner instability. In the six-month test, the long-term drift is included as well as variations in the phantom and coil positioning. Also, environmental factors may vary considerably. Multiple consecutive images can be used to define initial abnormal QA image detection thresholds for upcoming daily QA tests. A similar analysis of variations in consecutive measurements in multiple scanners is presented by Colombo et al. [7] with comparable results.

An important part of any QA is the communication of the results between participating groups including QA specialists, the users performing the tests, and service personnel. The in-house communication can be improved by a web-based results browser showing the key findings for all the scanners. The system may also be combined with automatic error detection with notifications when an abnormal test result is detected. This may improve the possibility of detecting an abnormal behavior before the fault affects the clinical image quality. Preventive and planned maintenance triggered by an abnormal result could potentially reduce scanner downtimes and limit appointment cancellations due to device malfunctions.

The limited number of scanners in the study does not allow a statistical comparison of characteristic differences between scanner types. For example, it would be expected that there is a systematic bias in QA parameters with respect to the scanner field strength.

Some limitations in the imaging process are introduced to achieve better applicability in clinical use. The scanning of the phantom was typically carried out within minutes from the table movement. Thus, the phantom fluid movement may not be completely stabilized before the scanning starts. This probably hinders the repeatability of the intensity uniformity measurements. Additionally, although the phantom position was fixed with a help of a phantom holder and the slice position relative to the coil was agreed upon, different operators carried out the measurements introducing variations to the measurements.

The slice width in the daily QA imaging sequence was 5 mm, which can be considered high compared with typical clinical imaging sequences. The relatively thick slice guarantees a high SNR, which is useful in an image processing pipeline where the phantom location needs to be detected accurately. Additionally, the accuracy of the ghosting detection increases when the relative amplitude of the background noise is small.

Image normalization was enabled in the QA imaging protocol. This may limit the sensitivity of the SNR and uniformity measurements. However, in the QA protocol the daily images were primarily used in a visual verification and the uniform image appearance was considered beneficial. Also, the normalization was enabled to be in line with the clinical scan protocols: noticeable deviations in the phantom images were postulated to be indicative of a potentially significant deterioration of patient images.

The phantoms used in this study were the standard QA phantoms provided by the manufacturers. Thus, the phantom diameters, shapes and rigidness varied from scanner to scanner. The measured parameters were relative in nature and the repeatability of the phantom positioning is an important requirement. However, some phantom shapes are more sensitive to variations in positioning than others: a spherical phantom or one with flexible casing is likely to produce variations at the phantom edges. This effect was suppressed by using a circular signal ROI with 80% of the signal producing area diameter. On the other hand, this may slightly decrease the sensitivity of the image intensity uniformity measurement, especially if the phantom diameter is relatively small.

5 Conclusions

The variations and baseline levels of image QA parameters can differ considerably between MRI scanner installations. Error thresholds for daily QA should be set individually for each scanner, because the results are affected by the exact hardware-phantom combination, QA imaging sequence and scanner environment. A clear dependence on the stability with respect to the scanner field strength or mobility was not found. Scanner specific thresholds based on image QA parameter means and standard deviations are a viable option to detect abnormalities in QA images.

Disclosures

The Authors declare that there are no conflicts of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements

We thank all the physicists and unit staff who has been participating in the QA program over the years.

References

1. Koller C, Eatough J, Mountford P et al. A survey of MRI quality assurance programmes. The British Journal of Radiology 2006;79:592-596.
2. American College of Radiology*:* MRI Quality Control Manual; 2015.
3. Fransson A, IPEM Publication, Report No. 80 — Quality control in Magnetic Resonance Imaging, R. Lerski, J. de Wilde, D. Boyce, J. Ridgway, Institute of Physics and Engineering in Medicine, UK; 1999.
4. International Engineering Consortium. IEC 62464-1 Magnetic Resonance Equipment for Medical Imaging—Part 1: Determination of Essential Image Quality Parameters; 2007.
5. National Electrical Manufacturers Association. NEMA Standards Publication MS 1-2008 Determination of Signal-to-Noise Ratio (SNR) in Diagnostic Magnetic Resonance Imaging; 2008.
6. National Electrical Manufacturers Association. NEMA Standards Publication MS 3-2008 Determination of Image Uniformity in Diagnostic Magnetic Resonance Images; 2008.
7. Colombo P, Baldassarri A, Del Corona M et al. Multicenter trial for the set-up of a MRI quality assurance programme. Magnetic Resonance Imaging 2004;22:93-101.
8. Peltonen JI, Mäkelä T, Sofiev A et al. An Automatic Image Processing Workflow for Daily Magnetic Resonance Imaging Quality Assurance. Journal of Digital Imaging 2017;30:163-171.
9. Ihalainen TM, Lönnroth NT, Peltonen JI et al. MRI quality assurance using the ACR phantom in a multi-unit imaging center. Acta Oncologica 2011;50:966-972.
10. Ihalainen T, Sipilä O, Savolainen S. MRI quality control: six imagers studied using eleven unified image quality parameters. European Radiology 2004;14:1859-1865.
11. Gunter JL, Bernstein MA, Borowski BJ et al. Measurement of MRI scanner performance with the ADNI phantom. Medical Physics 2009;36:2193-2205.
12. Kaljuste D, Nigul M. Evaluation of the ACR MRI phantom for quality assurance tests of 1.5 T MRI scanners in Estonian hospitals. Proceedings of the Estonian Academy of Sciences 2014;63:240.
13. Sun J, Barnes M, Dowling J et al. An open source automatic quality assurance (OSAQA) tool for the ACR MRI phantom. Australasian Physical & Engineering Sciences in Medicine 2014;38:39-46.
14. Fu L, Fonov V, Pike B et al. Automated analysis of multi site MRI phantom data for the NIHPD project. International Conference on Medical Image Computing and Computer-Assisted Intervention 2006;144-151.
15. Davids M, Zöllner FG, Ruttorf M et al. IMAGEN Consortium: Fully-automated quality assurance in multi-center studies using MRI phantom measurements. Magn Reson Imaging 2014;32:771-780.

16. Panych LP, Chiou Jr G, Qin L et al. On replacing the manual measurement of ACR phantom images performed by MRI technologists with an automated measurement approach. Journal of Magnetic Resonance Imaging 2016;43:843-852.

17. Simmons A, Moore E, Williams SC. Quality control for functional magnetic resonance imaging using automated data analysis and Shewhart charting. Magnetic Resonance in Medicine 1999;41:1274-1278.

18. Bourel P, Gibon D, Coste E et al. Automatic quality assessment protocol for MRI equipment. Medical Physics 1999;26:2693-2700.

19. McRobbie D, Quest R. Effectiveness and relevance of MR acceptance testing: results of an 8 year audit. The British Journal of Radiology 2002;75:523-531.

20. Iglewicz B, Hoaglin D. How to detect and handle outliers. Asq Press, 1993.

21. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin. 1968;70:213.

22. Hafey C. Cornerstone. 2016. https://github.com/chafey/cornerstone.

23. Eichelberg M, Riesmeier J, Wilkens T. Ten years of medical imaging standardization and prototypical implementation: the DICOM standard and the OFFIS DICOM toolkit (DCMTK). International Society for Optics and Photonics, Medical Imaging 2004: PACS and Imaging Informatics 2004:57-68.

24. Vanderkam D. Dygraphs. 2015. http://dygraphs.com.

Figures

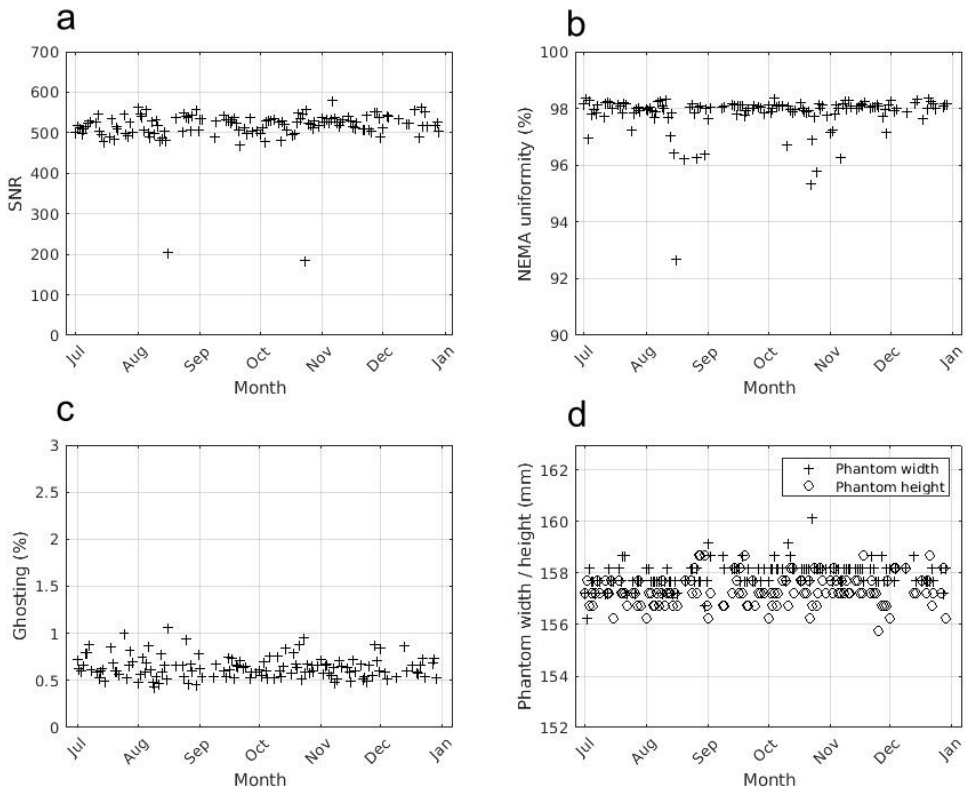

Fig. 1. Typical time series plots of a) the signal-to-noise ratio, b) image uniformity, c) image ghosting and d) phantom width and height. The outliers visible in images a) and b) resulted from faults in MRI system with no apparent reason. The scanner operated normally in the following scans without further actions.

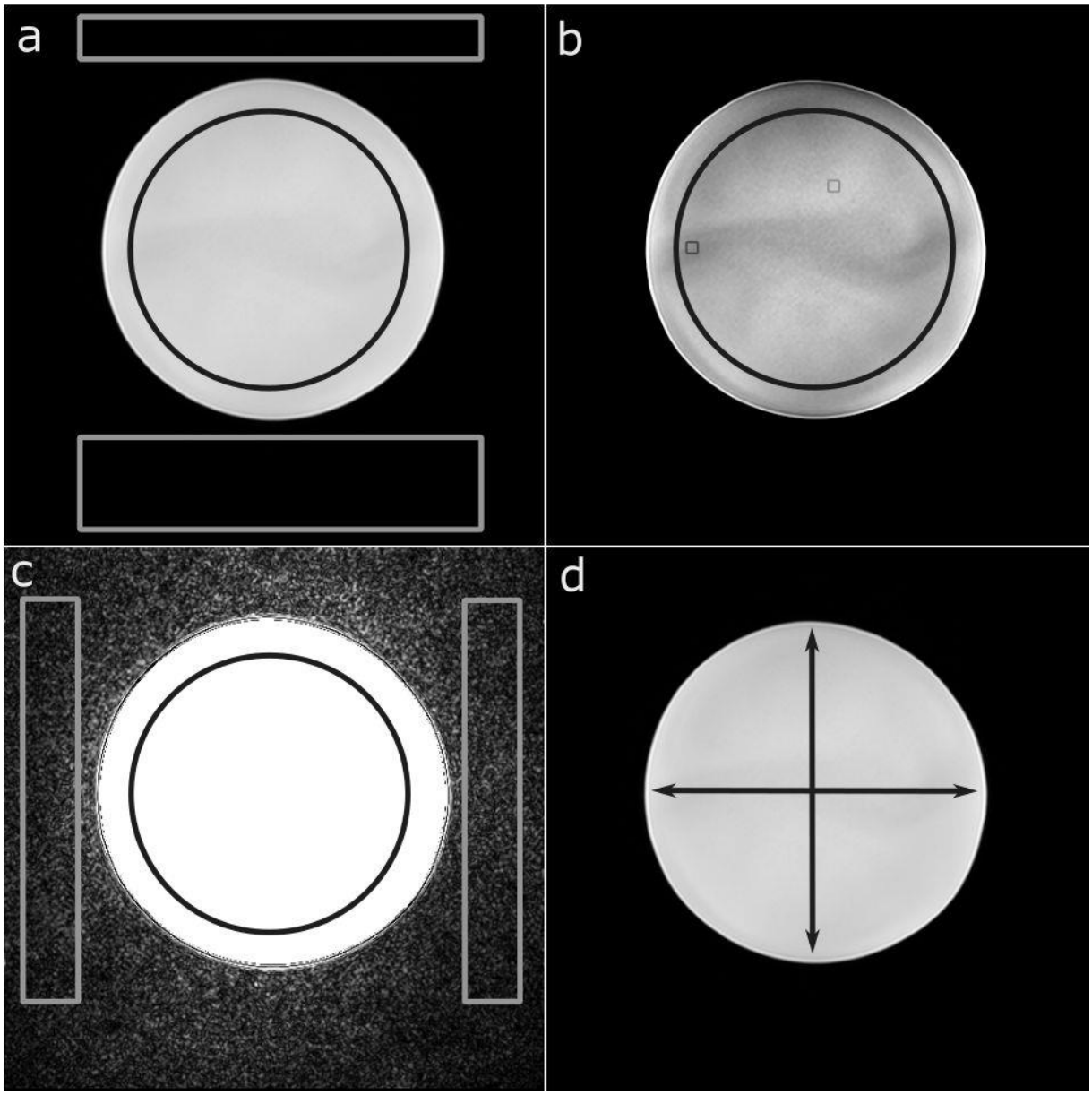

Fig. 2. An example of a typical daily quality assurance image and ROI placements in a) the signal-to-noise, b) image intensity uniformity, c) image ghosting and d) geometric distortion measurements.

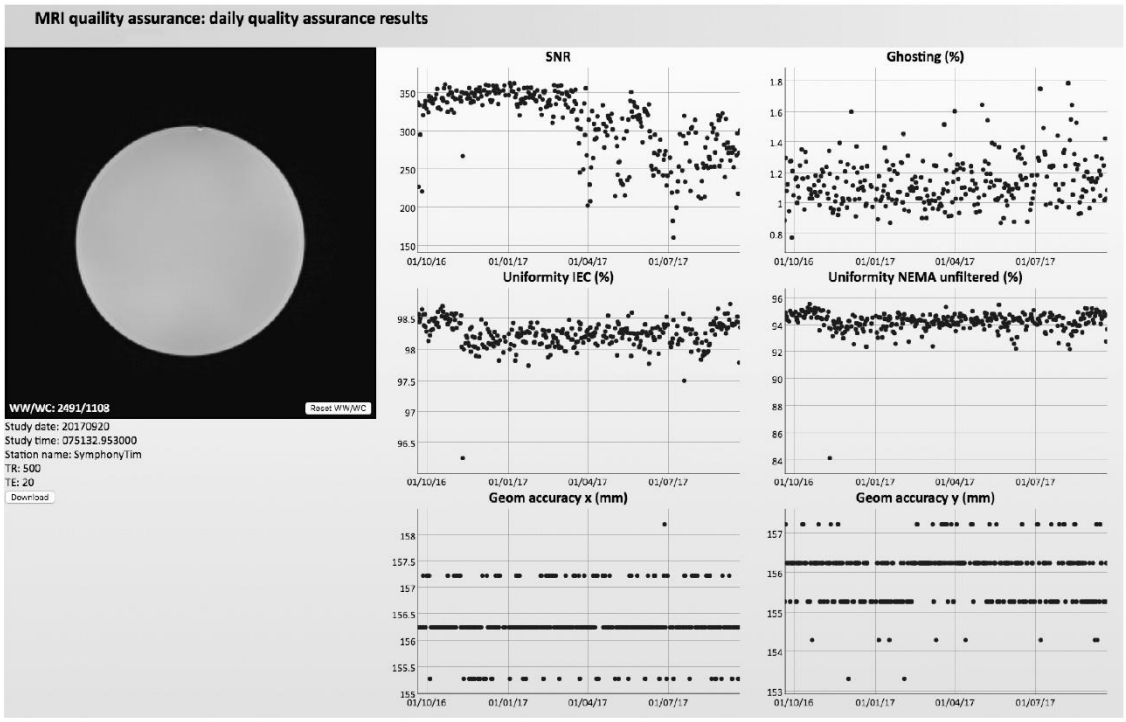

Fig. 3. The user interface to communicate daily QA results to the users.

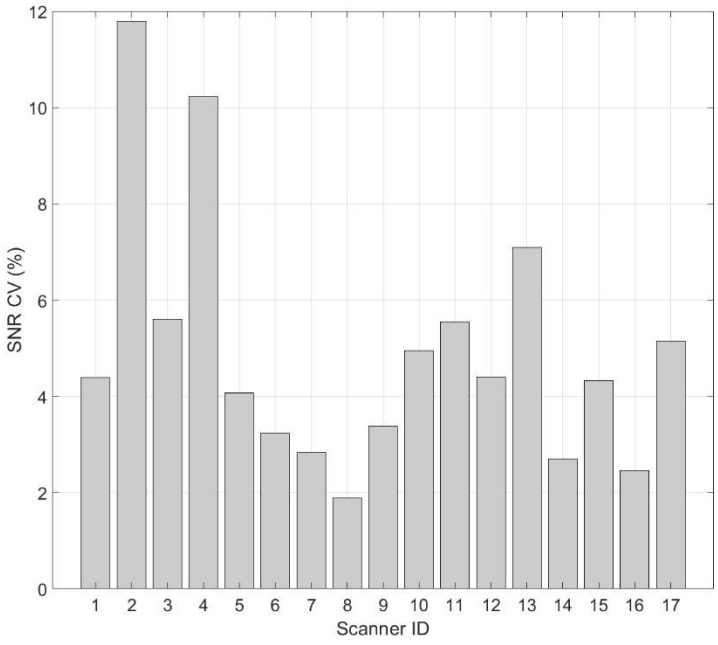

Fig. 4. The signal-to-noise ratio coefficients of variations for all the scanners.

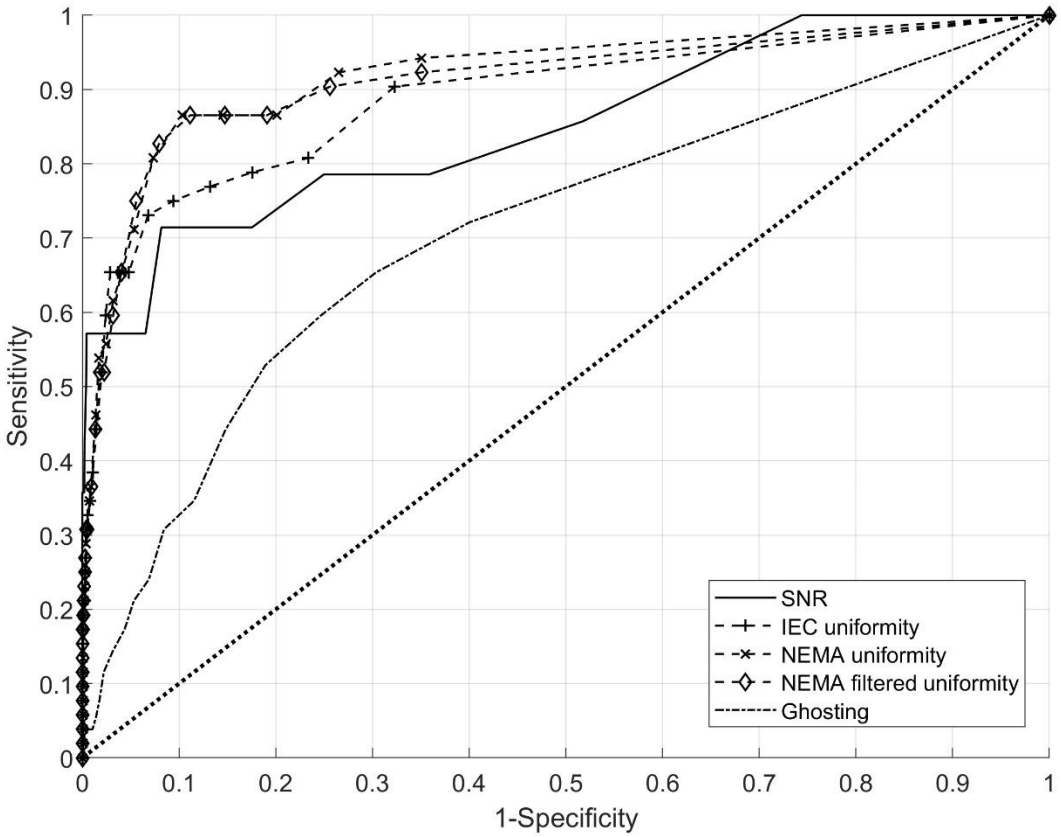

Fig. 5. The receiver operating characteristic curves of the quality assurance parameters. The detection thresholds were varied as standard deviations from the mean. Two-sided thresholds were used for the SNR and geometric distortions and one sided for the image ghosting and uniformity.

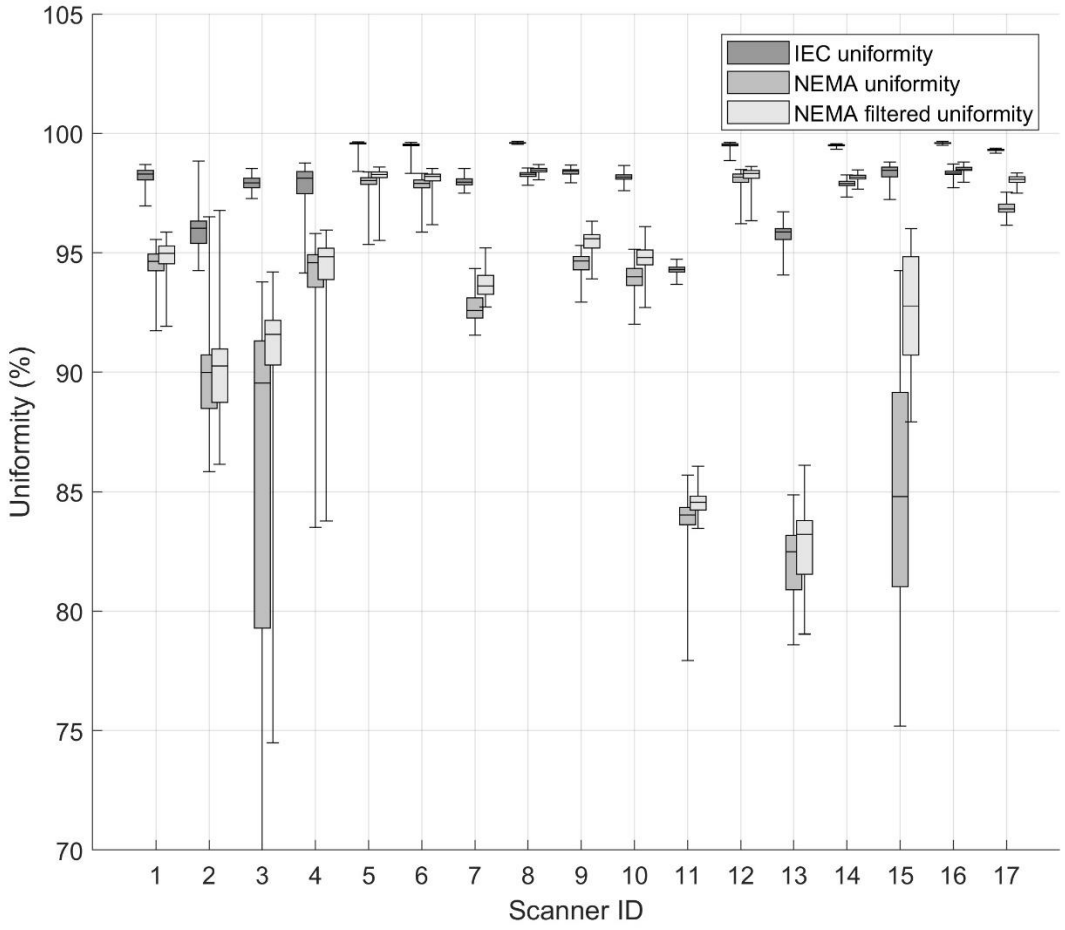

Fig. 6. Uniformity medians, ranges, and interquartile ranges for the three methods. The lowest outlier in the NEMA uniformity of scanner ID 3 (41.3 %) is outside the figure.

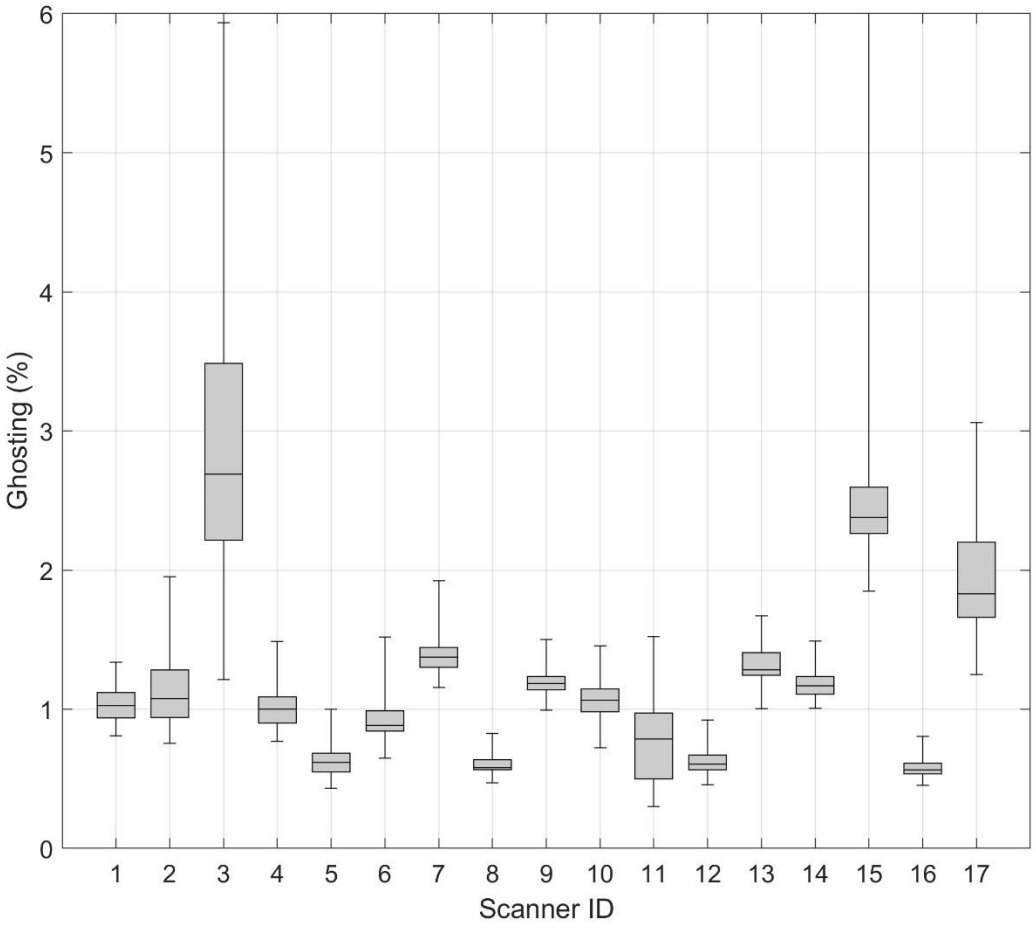

Fig. 7. Ghosting medians, ranges, and interquartile ranges. The highest ghosting (scanner ID 15, 10.9 %) is outside the figure.

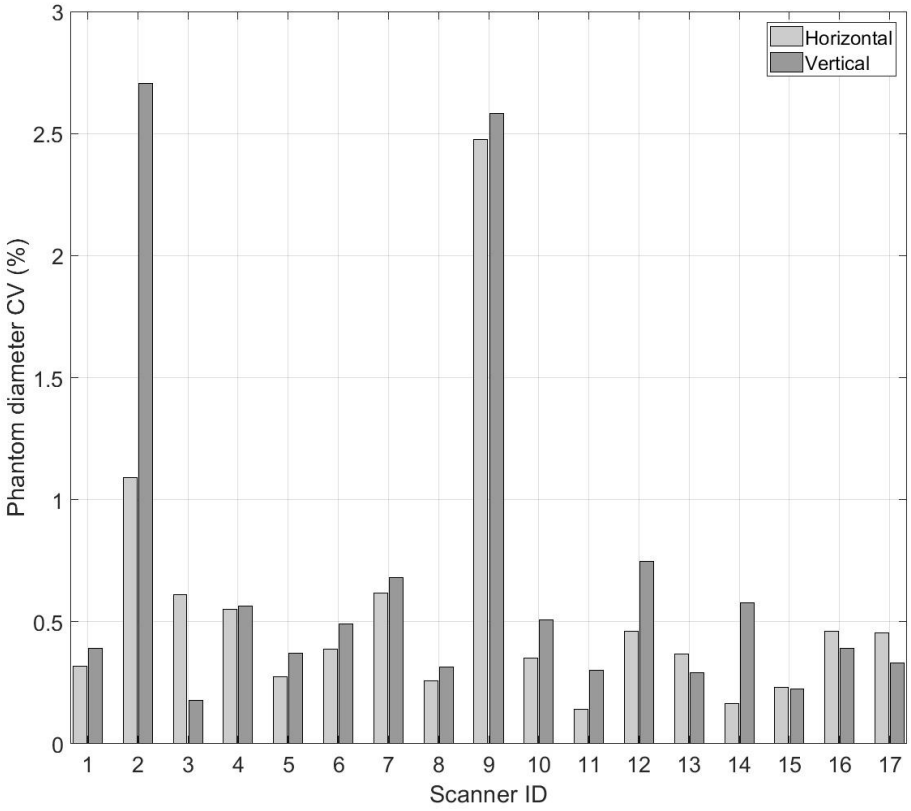

Fig. 8. The coefficient of variations of the measured phantom diameters in the horizontal and vertical directions.

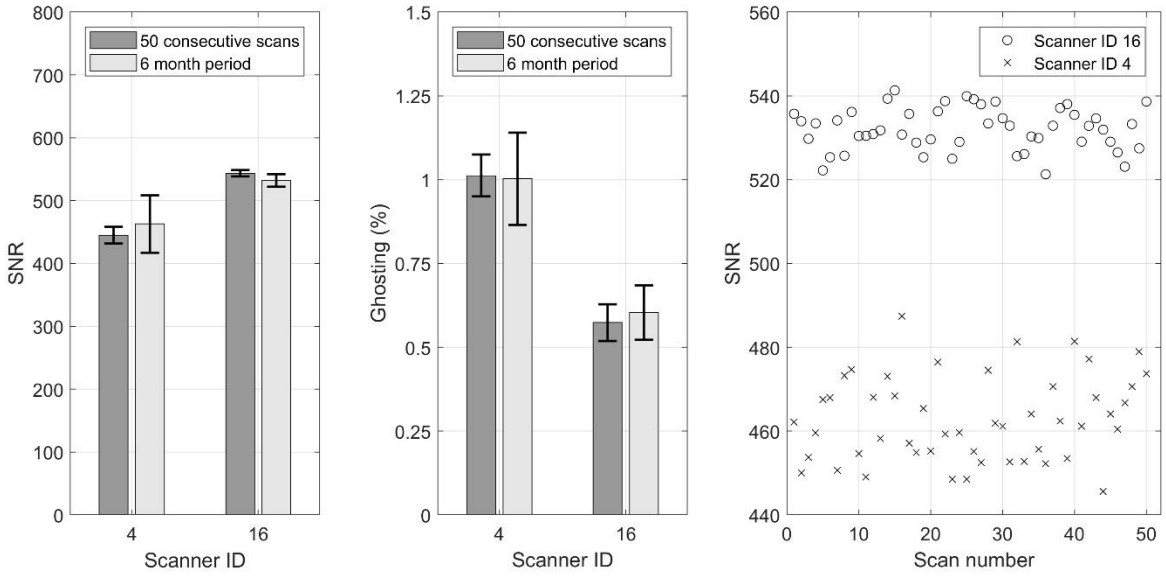

Fig. 9. A comparison between the signal-to-noise ratio (left) and ghosting (middle) results obtained with consecutive scans and results obtained during a 6-month test period with single SD whiskers. Signal-to-noise scatter plot of 50 consecutive scan in one session on the right.

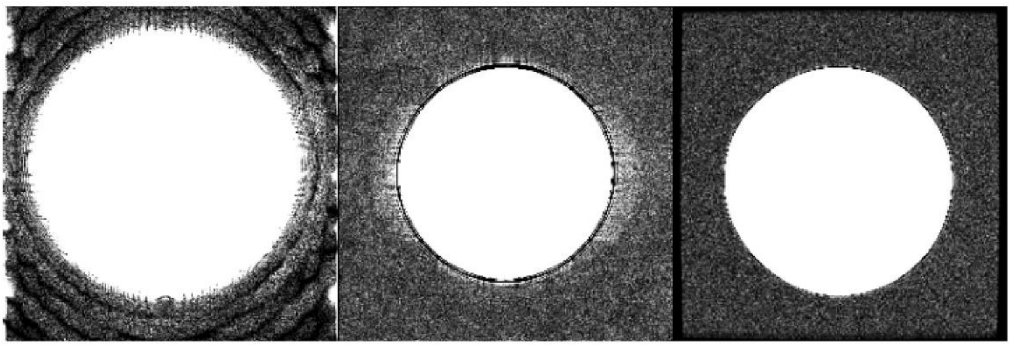

Fig. 10. Three examples of typical noise profiles in daily QA images from three vendors.

Biographies

Juha Peltonen works as a Medical Physicist in HUS Medical Imaging Center, Finland. His professional and research interests are specifically on medical image processing, quality assurance in radiology and magnetic resonance imaging. He is also a responsible physicist in the digital x-ray imaging in the HUS Medical Imaging Center and Helsinki University Central Hospital. Mr Peltonen is the president of the physicist detachment in the Finnish Radiology Association.

Lauri Lehmonen is a specializing physicist at HUS Medical Imaging Center, Finland. He is currently finalizing his training in the department of nuclear medicine. He's close to finishing his PhD involving motion quantification in cardiovascular magnetic resonance imaging.

Teemu Mäkelä works as a Medical Physicist at the Radiology Department of HUS Medical Imaging Center, Finland. His primary academic interests concern medical image processing, image quality measurements and machine learning. He is doing his PhD on convolutional neural network based analysis of computed tomography images.

Eero Salli is senior researcher at HUS Medical Imaging Center, Finland. He received his PhD degree in engineering physics from Helsinki University of Technology (currently Aalto University) in Finland in 2002. His research interests include medical image analysis and machine learning.

Biographies for the other authors are not available

Appendix

Table 2. Comparison of image quality parameter means and SDs between six-month test period and 50 consecutive scans. The latter were acquired in a single session.

	Signal-to-noise ratio		**Uniformity IEC (%)**		**Uniformity NEMA (%)**		**Uniformity NEMA filtered (%)**		**Ghosting (%)**		**Geometric distortion vert. (mm)**		**Geometric distortion horiz. (mm)**	
ID4 50 cons. scans mean (SD)	462,86	(10,14)	98,45	(0,01)	95,21	(0,16)	95,46	(0,12)	1,00	(0,08)	134,77	(0,00)	133,18	(0,48)
ID4 6 months mean (SD)	445,09	(45,51)	97,68	(1,13)	93,39	(2,84)	93,67	(2,85)	1,01	(0,14)	133,92	(0,74)	133,43	(0,75)
ID16 50 cons. scans mean (SD)	532,00	(5,05)	99,56	(0,01)	98,16	(0,09)	98,35	(0,08)	0,60	(0,05)	133,79	(0,00)	133,44	(0,22)
ID16 6 months mean (SD)	543,33	(13,36)	99,59	(0,02)	98,33	(0,14)	98,51	(0,12)	0,57	(0,06)	134,27	(0,62)	133,56	(0,52)

Table 3. Means and SDs (in parentheses) of image quality parameters.

Scanner ID	Signal-to-noise ratio	Uniformity IEC (%)	Uniformity NEMA (%)	Uniformity NEMA filtered (%)	Ghosting (%)	Geometric distortion horiz. (mm)	Geometric distortion vert. (mm)
1	409,4 (18,0)	98,2 (0,4)	94,5 (0,6)	94,9 (0,7)	1,0 (0,1)	132,0 (0,4)	132,0 (0,5)
2	345,0 (40,7)	95,9 (0,7)	89,8 (1,7)	90,0 (1,7)	1,1 (0,2)	157,2 (1,7)	153,5 (4,2)
3	232,2 (13,0)	97,9 (0,3)	85,6 (9,1)	90,9 (2,5)	2,9 (1,0)	199,0 (1,2)	197,3 (0,4)
4	445,1 (45,5)	97,7 (1,1)	93,4 (2,8)	93,7 (2,8)	1,0 (0,1)	133,9 (0,7)	133,4 (0,8)
5	521,4 (21,2)	99,5 (0,2)	97,9 (0,5)	98,1 (0,5)	0,6 (0,1)	158,0 (0,4)	157,3 (0,6)
6	389,5 (12,6)	99,5 (0,2)	97,8 (0,5)	98,1 (0,5)	0,9 (0,1)	157,8 (0,6)	156,9 (0,8)
7	153,9 (4,4)	98,0 (0,2)	92,7 (0,6)	93,7 (0,5)	1,4 (0,1)	167,1 (1,0)	167,3 (1,1)
8	543,4 (10,3)	99,6 (0,0)	98,3 (0,1)	98,5 (0,1)	0,6 (0,1)	131,8 (0,3)	132,2 (0,4)
9	170,6 (5,8)	98,4 (0,1)	94,5 (0,4)	95,5 (0,4)	1,2 (0,1)	162,1 (4,0)	161,9 (4,2)
10	340,0 (16,8)	98,2 (0,2)	93,9 (0,6)	94,7 (0,6)	1,1 (0,1)	156,1 (0,5)	155,9 (0,8)
11	501,5 (27,8)	94,3 (0,2)	83,9 (1,0)	84,5 (0,4)	0,7 (0,3)	199,0 (0,3)	197,8 (0,6)
12	587,0 (25,9)	99,5 (0,1)	98,1 (0,4)	98,2 (0,4)	0,6 (0,1)	157,5 (0,7)	157,1 (1,2)
13	216,4 (15,4)	95,7 (0,4)	82,1 (1,4)	82,8 (1,5)	1,3 (0,1)	158,3 (0,6)	156,0 (0,5)
14	350,4 (9,4)	99,5 (0,0)	97,9 (0,2)	98,2 (0,1)	1,2 (0,1)	157,3 (0,3)	153,9 (0,9)
15	272,5 (11,8)	98,4 (0,3)	85,2 (5,2)	92,7 (2,2)	2,5 (0,8)	198,9 (0,5)	199,1 (0,4)
16	543,3 (13,4)	99,6 (0,0)	98,3 (0,1)	98,5 (0,1)	0,6 (0,1)	134,3 (0,6)	133,6 (0,5)
17	139,5 (7,2)	99,3 (0,0)	96,9 (0,3)	98,1 (0,2)	2,0 (0,4)	89,4 (0,4)	88,3 (0,3)
Mean		**98,2 (0,3)**	**93,0 (1,5)**	**94,2 (0,9)**	**1,2 (0,2)**		

Table 4. Medians and modified Z-scores (in parenthesis) of image quality parameters. Z-scores under -3.5 and over 3.5 are bolded.

Scanner ID	Signal-to-noise ratio	Uniformity IEC (%)	Uniformity NEMA (%)	Uniformity NEMA filtered (%)	Ghosting (%)	Geometric distortion horiz. (mm)	Geometric distortion vert. (mm)
1	410,6	98,3 (-0,08)	94,6 (+0,00)	95,0 (+0,00)	1,0 (-0,10)	131,8	131,8
2	343,3	96,0 (-1,48)	90,0 (-0,93)	90,3 (-0,99)	1,1 (+0,02)	157,0	155,3
3	235,0	97,9 (-0,30)	89,5 (-1,02)	91,6 (-0,71)	2,7 **(+3,92)**	198,2	197,3
4	458,3	98,1 (-0,18)	94,6 (-0,01)	94,8 (-0,03)	1,0 (-0,16)	133,8	133,8
5	524,7	99,6 (+0,72)	98,0 (+0,67)	98,3 (+0,69)	0,6 (-1,08)	158,2	157,2
6	387,7	99,5 (+0,69)	97,9 (+0,65)	98,2 (+0,67)	0,9 (-0,44)	157,7	156,7
7	153,4	98,0 (-0,29)	92,6 (-0,41)	93,6 (-0,29)	1,4 (+0,75)	167,0	167,0
8	542,6	99,6 (+0,73)	98,3 (+0,73)	98,5 (+0,73)	0,6 (-1,17)	131,8	132,3
9	169,8	98,4 (+0,00)	94,7 (+0,00)	95,6 (+0,13)	1,2 (+0,29)	160,2	160,2
10	342,8	98,2 (-0,15)	94,0 (-0,13)	94,8 (-0,04)	1,1 (+0,00)	156,3	156,3
11	498,9	94,3 (-2,56)	84,0 (-2,12)	84,6 (-2,18)	0,8 (-0,67)	199,0	198,0
12	589,8	99,5 (+0,69)	98,2 (+0,70)	98,3 (+0,70)	0,6 (-1,11)	157,7	156,7
13	218,9	95,9 (-1,58)	82,5 (-2,42)	83,2 (-2,46)	1,3 (+0,52)	158,0	156,0
14	349,6	99,5 (+0,67)	97,9 (+0,65)	98,2 (+0,67)	1,2 (+0,25)	157,2	153,8
15	273,8	98,5 (+0,02)	84,8 (-1,96)	92,8 (-0,46)	2,4 (+3,17)	199,2	199,2
16	543,6	99,6 (+0,73)	98,3 (+0,74)	98,5 (+0,74)	0,6 (-1,21)	134,3	133,8
17	138,2	99,3 (+0,55)	96,8 (+0,44)	98,1 (+0,65)	1,8 (+1,84)	89,1	88,6
Median		**98,4**	**94,6**	**95,0**	**1,1**		

Table 5. CVs and modified Z-scores (in parentheses) of image quality parameters. Z-scores under -3.5 and over 3.5 are bolded.

Scanner ID	Signal-to-noise ratio	Uniformity IEC (%)	Uniformity NEMA (%)	Uniformity NEMA filtered (%)	Ghosting (%)	Geometric distortion horiz. (mm)	Geometric distortion vert. (mm)
1	4,4 % (+0,00)	0,4 % (+1,54)	0,7 % (+0,09)	0,7 % (+0,28)	10,8 % (-0,46)	0,3 % (-0,37)	0,4 % (-0,00)
2	11,8 % **(+4,28)**	0,8 % **(+4,88)**	1,9 % (+1,93)	1,9 % (+2,52)	21,7 % (+1,49)	1,1 % **(+3,66)**	2,7 % **(+13,40)**
3	5,6 % (+0,70)	0,3 % (+0,67)	10,7 % **(+14,93)**	2,8 % **(+4,20)**	33,5 % **(+3,60)**	0,6 % (+1,15)	0,2 % (-1,24)
4	10,2 % **(+3,38)**	1,2 % **(+8,29)**	3,0 % **(+3,60)**	3,0 % **(+4,65)**	13,6 % (+0,04)	0,5 % (+0,84)	0,6 % (+1,00)
5	4,1 % (-0,19)	0,2 % (+0,34)	0,5 % (-0,14)	0,5 % (+0,00)	17,2 % (+0,68)	0,3 % (-0,60)	0,4 % (-0,12)
6	3,2 % (-0,67)	0,2 % (+0,31)	0,5 % (-0,19)	0,5 % (-0,08)	15,8 % (+0,44)	0,4 % (+0,00)	0,5 % (+0,56)
7	2,8 % (-0,90)	0,2 % (+0,00)	0,6 % (+0,00)	0,6 % (+0,07)	10,2 % (-0,57)	0,6 % (+1,20)	0,7 % (+1,67)
8	1,9 % (-1,45)	0,0 % (-1,49)	0,1 % (-0,73)	0,1 % (-0,78)	10,6 % (-0,48)	0,3 % (-0,67)	0,3 % (-0,45)
9	3,4 % (-0,59)	0,1 % (-0,54)	0,4 % (-0,29)	0,4 % (-0,19)	7,0 % (-1,13)	2,5 % **(+10,85)**	2,6 % **(+12,69)**
10	4,9 % (+0,32)	0,2 % (-0,08)	0,6 % (+0,00)	0,6 % (+0,15)	11,6 % (-0,31)	0,4 % (-0,20)	0,5 % (+0,67)
11	5,6 % (+0,67)	0,2 % (-0,10)	1,2 % (+0,85)	0,5 % (-0,01)	35,9 % **(+4,03)**	0,1 % (-1,29)	0,3 % (-0,53)
12	4,4 % (+0,01)	0,1 % (-0,54)	0,4 % (-0,34)	0,4 % (-0,27)	13,4 % (+0,00)	0,5 % (+0,38)	0,7 % (+2,05)
13	7,1 % (+1,57)	0,4 % (+2,20)	1,8 % (+1,70)	1,8 % (+2,27)	9,6 % (-0,67)	0,4 % (-0,11)	0,3 % (-0,58)
14	2,7 % (-0,98)	0,0 % (-1,34)	0,2 % (-0,67)	0,1 % (-0,73)	8,8 % (-0,82)	0,2 % (-1,16)	0,6 % (+1,08)
15	4,3 % (-0,04)	0,3 % (+0,60)	6,1 % **(+8,07)**	2,4 % (+3,42)	31,1 % (+3,16)	0,2 % (-0,82)	0,2 % (-0,97)
16	2,5 % (-1,12)	0,0 % (-1,43)	0,1 % (-0,71)	0,1 % (-0,76)	10,8 % (-0,45)	0,5 % (+0,37)	0,4 % (+0,00)
17	5,1 % (+0,44)	0,0 % (-1,32)	0,3 % (-0,53)	0,2 % (-0,67)	22,3 % (+1,59)	0,5 % (+0,34)	0,3 % (-0,35)
Mean	**4,9 %**	**0,3 %**	**1,7 %**	**1,0 %**	**16,7 %**	**0,5 %**	**0,7 %**

Table 6. Combined mean QA parameter CVs based on scanner field strength and mobility.

	Signal-to-noise ratio	**Uniformity IEC (%)**	**Uniformity NEMA (%)**	**Uniformity NEMA filtered (%)**	**Ghosting (%)**	**Geometric distortion vert. (mm)**	**Geometric distortion horiz. (mm)**
3T	8,0 %	0,6 %	4,1 %	2,1 %	19,9 %	0,6 %	1,0 %
1.5T static	4,1 %	0,2 %	1,2 %	0,7 %	15,4 %	0,5 %	0,7 %
1.5T mobile	3,4 %	0,2 %	0,5 %	0,5 %	14,4 %	0,4 %	0,5 %