CrossMark

MRI quality assurance based on 3D FLAIR brain images

Juha I. Peltonen[1,2] · Teemu Mäkelä[1,3] · Eero Salli[1]

Abstract

Objective Quality assurance (QA) of magnetic resonance imaging (MRI) often relies on imaging phantoms with suitable structures and uniform regions. However, the connection between phantom measurements and actual clinical image quality is ambiguous. Thus, it is desirable to measure objective image quality directly from clinical images.

Materials and methods In this work, four measurements suitable for clinical image QA were presented: image resolution, contrast-to-noise ratio, quality index and bias index. The methods were applied to a large cohort of clinical 3D FLAIR volumes over a test period of 9.5 months. The results were compared with phantom QA. Additionally, the effect of patient movement on the presented measures was studied.

Results A connection between the presented clinical QA methods and scanner performance was observed: the values reacted to MRI equipment breakdowns that occurred during the study period. No apparent correlation with phantom QA results was found. The patient movement was found to have a significant effect on the resolution and contrast-to-noise ratio values.

Discussion QA based on clinical images provides a direct method for following MRI scanner performance. The methods could be used to detect problems, and potentially reduce scanner downtime. Furthermore, with the presented methodologies comparisons could be made between different sequences and imaging settings. In the future, an online QA system could recognize insufficient image quality and suggest an immediate re-scan.

Keywords Magnetic resonance imaging · Quality assurance · Quality control · Computer-assisted image analysis

Introduction

Image quality assurance (QA) in magnetic resonance imaging (MRI) is often based on phantom tests defined in standards and guidelines [1–5] or by the manufacturer. Quantitative phantom measurements can characterize some aspect of the scanner's absolute imaging performance, but the relationship with the actual clinical image quality is often unclear. The selected imaging sequences may emphasize effects not observed with other techniques. Phantom images are often acquired with robust 2D (e.g. conventional spin echo) sequences which are prone to different characteristic artefacts than 3D sequences [6]. Additionally, human anatomy provides exceedingly more complex imaging target, including non-voluntary movement and flow. Thus, the scanner performance cannot be entirely predicted by phantom studies alone.

In addition to phantom based QA, it would be rational to measure image quality directly from the clinical images. However, the clinical image quality assessment is mostly based on qualitative observer-based ranking in the Likert scale [7] or similar. This approach is susceptible to intra- and interobserver variation and, therefore, lack reproducibility. The grading criteria can differ between (or even within) departments [8]. Quantitative computational methods for clinical image QA would enable the clinically relevant and reproducible assessment of MRI hardware and imaging sequence performance. It would also offer a possibility to compare scanners and sequences in a uniform scale.

There are only a few published quantitative methods designed to analyze clinical images in the sense of image

✉ Juha I. Peltonen
  juha.peltonen@hus.fi

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

2   Department of Neuroscience and Biomedical Engineering,
    School of Science, Aalto University, P.O. Box 12200,
    FI-00076 Espoo, Finland

3   Department of Physics, University of Helsinki, P.O. Box 64,
    FI-00014 Helsinki, Finland

quality [9–14]. Magnotta et al. presented a method for calculating signal-to-noise and contrast-to-noise (CNR) from clinical 2D images. Weng-Tung et al. studied 2D image resolution by applying radiofrequency tagging to images. Mortamet et al. presented two methods using an image volume background to derive QC measures. More recently, Borri et al. have used image power spectrum analysis to assess image spatial resolution, Osadebey et al. have developed quality measures based on local entropy in the images and Jang et al. have used feature statistics to track distortions in the images. A large cohort of clinical images has been analysed in respect for image quality in studies by Gedamu et al. and Kruggel et al. [15, 16]. A proposal for a complete pipeline capable of automatic image analysis was presented by Gedamu [17]. Additional methods for the detection of motion artefacts were presented in studies by Gedamu et al., Backhausen et al. and White et al. [18–20]. These methods measured also general image quality although their subject had a very specific error source.

In this work, we present four methods of assessing the image quality of 3D fluid attenuation inversion recovery (FLAIR) MRI sequence from clinical brain images. These methods were bundled as a novel automated pipeline and applied to a large cohort of clinical brain studies. The pipeline was utilized to demonstrate variations and trends in MRI scanner performance to assess variations in the scanner stability in long and short term. The results obtained with clinical volumes were compared with phantom QA results. Additionally, the effect of motion artefacts in the presented methods was studied.

Materials and methods

Imaging sequence

FLAIR is a valuable MRI technique for the detection of intracranial hemorrhage [21, 22]. The 3D FLAIR sequence used in this study was a turbo spin echo-based sequence involving radio frequency (RF) inversion pre-pulse and variable angle refocusing pulses to optimize contrast in the image [23]. The 3D sequence is nowadays often applied in brain studies instead of more traditional 2D sequences to decrease the duration of imaging protocol while maintaining adequate contrast between brain tissues and increasing imaging resolution. 3D FLAIR sequences are also less prone to cerebrospinal fluid (CSF) flow artefacts than their 2D counterparts [6]. In our department, the 3D FLAIR sequence is the most common imaging sequence used in brain studies.

The sequence can be optimized to produce optimal signal-to-noise ratio or modulation transfer function (MTF). These properties are somewhat interconnected and are only partly adjustable by the user [24]. Thus, the effect of changes

Table 1 Sequence parameters of the clinical 3D FLAIR sequence

Parameter	Value
Sequence type	3D FLAIR
Acquisition time	6 min 56 s
TE	376 ms
TR	8000 ms
TI	2400 ms
FOV (FH×AP×RL)	240×240×180 (mm)
Acquisition voxel size (FH×AP×RL)	1.0×1.1×1.2 (mm)
Reconstruction voxel size (FH×AP×RL)	1.0×1.0×1.0 (mm)
Echo train length	125
Frequency encoding direction	R–L
Bandwidth	521 Hz/pxl
Parallel imaging factor	4 (SENSE)

TE echo time, *TR* repetition time, *TI* inversion time, *FH* feet–head, *AP* anterior–posterior, *RL* right–left

in parameters and the scanner performance on the image quality are not entirely predictable. In our study, we used both phantom and clinical head volumes scanned with the 3D FLAIR sequence on a 3 T MRI scanner. The sequence parameters presented in Table 1 were used unless stated otherwise.

Image quality analysis

Preprocessing

As a first step of the analysis pipeline, the brain volume was extracted from the original image volume with the Statistical Parametric Mapping toolbox (SPM, http://www.fil.ion.ucl.ac.uk/spm/software/spm12). SPM generated brain tissue probability maps for white matter (WM), grey matter (GM), CSF and other tissue types. In addition, SPM produced a bias corrected version of the original image volume. The segmentation settings were SPM12 defaults (bias FWHM: 60 mm cutoff, bias regularization: 0.001, number of tissue types: 6). These maps were employed to generate an initial brain mask as a union of voxels which had at least 0.35 probability in WM or GM map (Fig. 1). From the initial brain mask, all but the largest connected object was removed, and remaining holes were filled to obtain the final brain mask.

In addition to the brain mask, a mask for the head volume was generated to separate the signal producing volume from the background (Fig. 2). First, the whole image volume was thresholded. The used thresholding level was derived semi-empirically by applying value obtained with Otsu's method [25] divided by four. After the thresholding, all but the largest volume was removed from the mask. Next, the remaining structure was dilated with a spherical structuring element
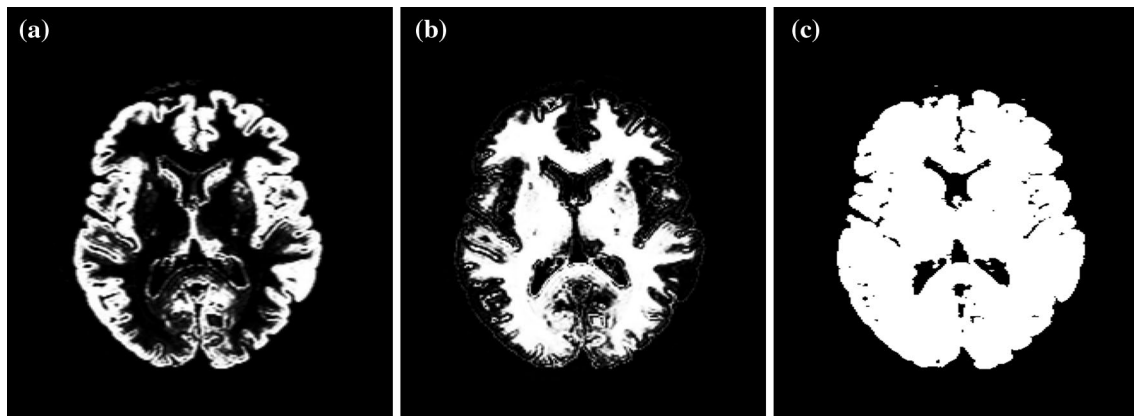
Fig. 1 An axial slice of typical probability maps for **a** WM, **b** GM and **c** initial brain mask

Fig. 2 **a** An illustration of segmented brain and head volumes. **b** Illustration of the outer perimeters of the head and brain mask in one transversal image slice
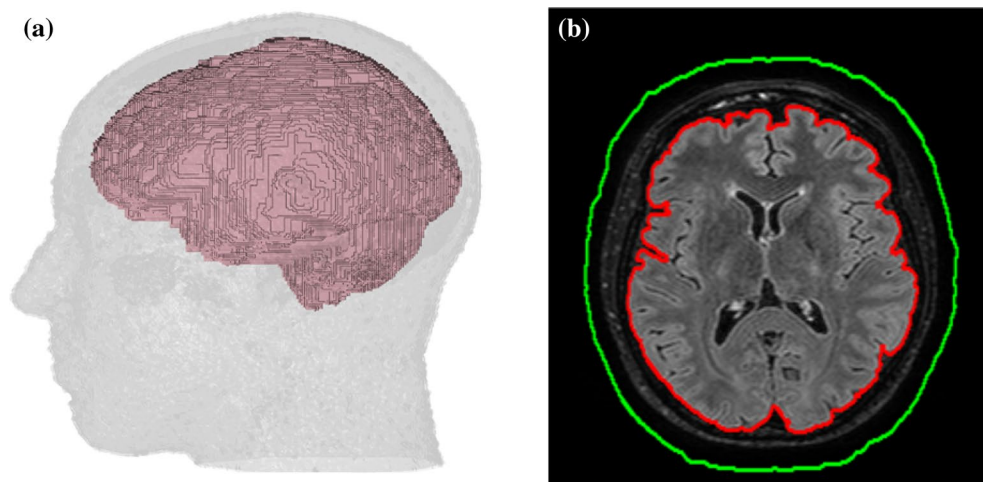
with the radius of 10 voxels which after remaining holes in the object were filled [26].

Resolution

The resolution of an imaging system describes its ability to reproduce sharp material interfaces and distinguish closely spaced features from each other. The former can be quantified by differentiating an edge spread function (ESF) and Fourier transforming resulting in line spread function (LSF) to obtain modulation transfer function (MTF). MTF describes an imaging system's spatial frequency contrast response. [27].

A well-defined edge is an essential requirement for MTF measurement. This condition can be easily satisfied using phantoms, but it becomes problematic when clinical images are assessed. In this study, we used the cortical surface as an edge for resolution measurement. The strong contrast and a sharp interface provide a favorable target surface.

In our method, the bias corrected image volume was first interpolated to isotropic $0.5 \times 0.5 \times 0.5$ mm^3 resolution.
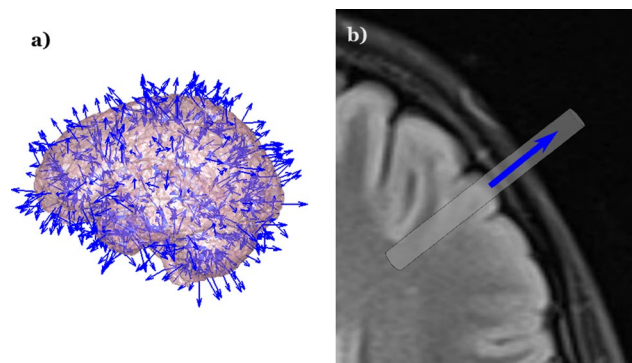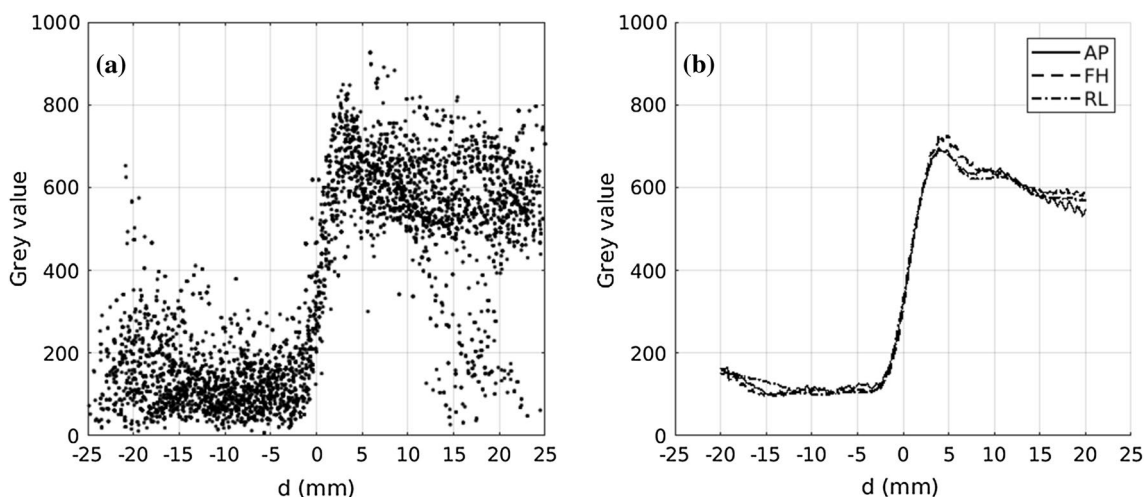

Fig. 3 **a** A sample of vectors perpendicular to the brain surface and **b** an illustration of the normal vector and corresponding cylinder (right)

Tetrahedral mesh was generated from the brain mask with iso2mesh library [28]. Typically, from 50,000 to 95,000 triangular polygons were generated from which 30,000 randomly selected were used in the resolution measurement. Each of these polygons was used to define a cylinder with a
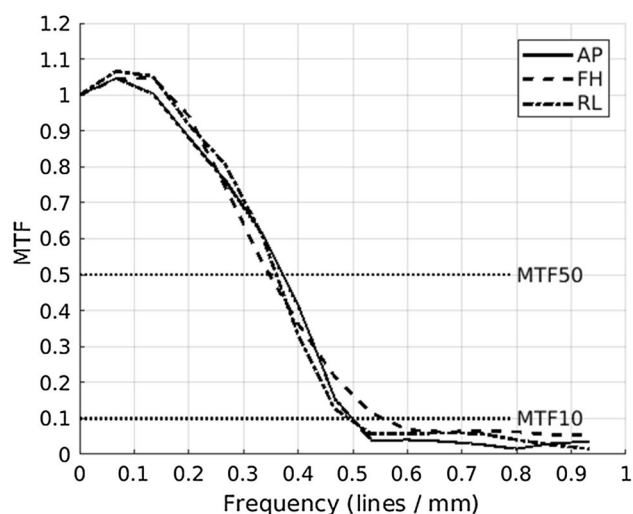
diameter of 1 mm and direction perpendicular to the brain surface (Fig. 3). The grey value and the distance from the mesh were recorded for the voxels inside the cylinder and used to create preliminary ESFs. The location of the edge in each preliminary ESF was refined by 1D convolution with the first derivative of Gaussian and finding the maximum [29]. Each preliminary profile origin was then shifted correspondingly to the center of the detected edge.

Each preliminary ESF was verified to represent an actual edge of the brain surface. The automatic verifying methods included a check that difference of the grey values over the edge is reasonable and to check that the 1D convolution resulted in high enough values to guarantee a reasonable gradient of the edge. Also, there was a limit on the maximum detected edge distance (5 mm) from the mesh. Filtered ESFs could then be averaged together at selected directions. In this work, we studied ESFs in three orthogonal directions corresponding to anatomical volume alignment: anterior–posterior (AP), feet–head (FH) and right–left (RL). The opening angle limiting accepted preliminary ESFs in each direction was 15°. An example of directional ESFs is presented in Fig. 4 together with point cloud representing a sample of all ESF points.

The averaged ESFs were then differentiated to obtain LSFs. Before Fourier transformation, the LSFs were Hanning filtered to suppress high frequency components originating from the LSF tails [27]. Typical resulting directional MTFs are presented in Fig. 5. The used MTF10 and MTF50 values were chosen as resolution measures corresponding spatial frequencies where the MTF is 10% and 50% of the value at zero spatial frequency as presented in Fig. 5.

Fig. 5 Typical directional MTFs in AP, FH and RL directions

Quality index

Instabilities during scanning can produce ambiguity in the image reconstruction. This can result, e.g. from the patient movement or imperfect operation of gradient fields. Signal intensity can then spill to erroneous locations inside the imaging volume. The most evident is the introduction of extra signal outside the actual signal producing volume. Thus, the amount of signal outside the anatomical volume can be used as a figure of merit. A quality index (QI) adapted in this study has been presented by Mortamet et al. [11], and it can be calculated by

$$QI = \frac{N_{\text{artefact}}}{N_{\text{BG}}} \tag{1}$$

Fig. 4 **a** An example of point cloud consisting of a sample of 5000 points taken from ESFs in AP direction. **b** Typical directional ESF profiles in AP, FH and RL directions

where N_{BG} is the number of background voxels and $N_{artefact}$ is the number of background voxels labelled as artefactual. Background voxels were defined as image volume excluding the head volume and a 10 voxel margin to the image volume borders. As proposed by Mortamet et al., voxels were labeled as artefactual by thresholding background volume with the value corresponding to the peak of the background histogram and eroding and dilating the result with 3D cross-structuring element [26].

Contrast-to-noise ratio

CNR is an image quality parameter that reflects the imaging systems ability to differentiate noise contaminated objects by their signal level in the image. The contrast can be defined as a relative difference in the signal strengths of two known objects in the image. The discernibility at different contrast levels is limited by the amount of noise. The CNR parameter in this study was calculated by

$$\text{CNR} = \frac{\text{GM} - \text{WM}}{\sigma_{BG}} \tag{2}$$

The representative WM and GM grey values were obtained from the model fitting parameters of the SPM software used for tissue segmentation. The software package is fitting a mixture Gaussian model to brain and the used WM and GM values are expectation values of Gaussian distributions respectively. σ_{BG} is the standard deviation of voxel grey values in original image volume excluding the head volume and 10 voxel margins to the image volume border.

Image intensity spatial homogeneity

Image intensity level inhomogeneity may be induced to MRI by spatial differences in a scanner's RF-transmit (B1) or a receive field. Modern MRI scanners have advanced methods of correcting signal inhomogeneities in the images. If there is substantial signal inhomogeneity present after these corrections, it may be a sign of a hardware failure.

The level of signal inhomogeneity was studied using the bias correction feature of the SPM package which produced an intensity corrected version of the original image volume. The average bias correction of a voxel, or bias index (BI), was then calculated by

$$\text{BI} = 100\% \times \frac{\sum_{i=1}^{N} \left| \frac{\text{bias corrected}_i - \text{original}_i}{\text{original}_i} \right|}{N} \tag{3}$$

where N is the number of voxels included in the spherical volume with a 10 cm diameter and concentric with the brain mask center of mass, bias corrected is the grey value in the bias corrected volume, original is the intensity in original volume and i is the voxel index.

Resolution measurement validation

The resolution calculation was validated with a standard spherical quality assurance phantom with the diameter of 17 cm and liquid signal producing content. In the testing, the phantom was scanned with variable isotropic voxel sizes of 0.9–1.3 mm with step of 0.2 mm. The imaging sequence was otherwise identical to the one used in the clinical images.

Additionally, the resolution calculation method was tested by filtering a binary brain mask with 3D Gaussian filter and studying how the parameters of the filter affect the resulting MTF10 and MTF50 values. The results were compared with simulated ideal MTF10 and MTF50 values obtained by the Gaussian filtration of a 1D step function.

Clinical head volume analysis

The presented image quality metrics was calculated for a large cohort of clinical head volumes spanning over a test period of 9.5 months. In total, 665 head volumes were included. The inclusion criterion of the head volume was a GM/WM volume ratio between 0.3 and 2.0 to filter out volumes with substantial pathologies and inaccurate segmentation. The inclusion criterion was verified visually. The GM and WM volumes were calculated as the sum of respective probability maps generated by SPM. Fifty-five percent of the patients were female, 45% male and the median age was 43 (range 13–84) years. The possible presence of foreign objects (such as metallic implants) was not taken into consideration in the analysis. The use of clinical volumes was approved by department's scientific committee.

The mean and standard deviation for each parameter were calculated for each day of the period by using a 7-day moving window. Each 7-day window included in average 19.4 studies with the standard deviation of 5.5 studies. During the test period, there were two major scanner breakdowns. The first breakdown occurred at the beginning of the month number three and was caused by a broken RF-amplifier. The second breakdown occurred at the beginning of the month number eight and was caused by a break in a gradient coil requiring the full replacement of the gradient coil system.

Comparison with phantom measurements

During the clinical QA test period, a daily phantom QA program was in position. A cross-sectional image of a manufacturer provided cylindrical phantom was scanned daily

using a head coil, fixed phantom position and a standard spin echo (SE) sequence. The primary purpose of acquiring the daily QA image was to verify the scanner was working properly before the first patient of the day. In addition to visual inspection, the image was sent to a QA server for detailed analysis. Calculated QA parameter time series included signal-to-noise ratio (SNR), image intensity uniformity, image ghosting and geometric distortion. SNR was calculated by a single image method presented by National Electrical Manufacturers Association (NEMA) [4], ghosting was calculated as presented by International Engineering Consortium (IEC) [3] and image intensity uniformity with methods presented by both IEC [3] and NEMA [5]. The geometric distortion was followed by measuring a phantom diameter in horizontal and vertical directions. A full description of the utilized automatic daily phantom QA pipeline is presented by Peltonen et al. [30].

The effect of motion artefact

The effect of patient motion artefact was studied by labeling all head volumes as normal (91%) or affected by patient motion (9%). The labeling was done by an experienced QA specialist (JIP) based on the amount of blurring in a central axial slice. The median, interquartile range and range of all image quality parameters were calculated over all image volumes for three cases: all the data, only the non-artefact volumes and only the artefact-positive volumes. The statistical difference between the images was studied with two-sample Student's t test.
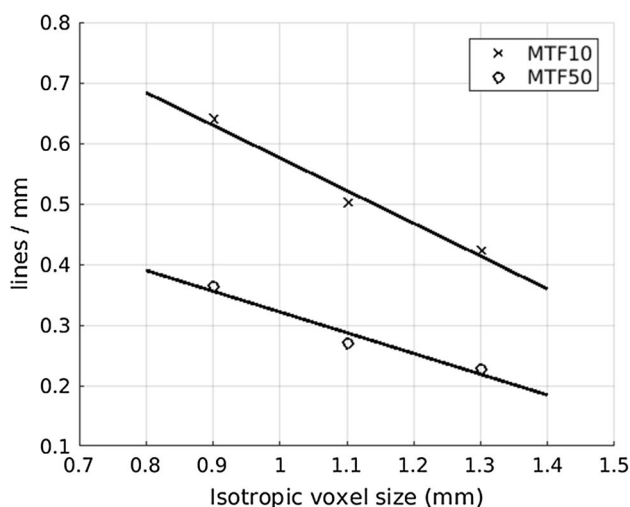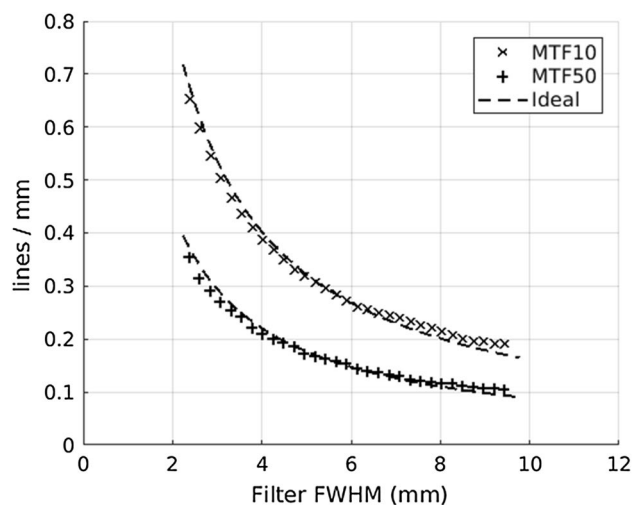
Results

The image resolution measurement method was validated by imaging a spherical phantom with variable voxel sizes. The effect on the voxel size to measured MTF10 and MTF50 values in the ball phantom is presented in Fig. 6. Voxel size has a linear relation to the measured resolution variables MTF10 ($R2 = 0.98$) and MTF50 ($R2 = 0.95$). Additionally, a test was done with a single brain mask (Fig. 7) where the effect of 3D Gaussian filtering on the measured resolution parameters was studied. The MTF10 and MTF50 value response to filtering was close to the ideal response.

The directional MTF10 and MTF50 values with 7-day running average and standard deviation measured from the clinical head scans during the 9.5-month period are presented in Fig. 8. In the FH direction there is a decrease of the resolution values before the both breakdowns of the scanner. Additionally, MTF10 and MTF50 values in FH direction are increased after the gradient system breakdown in month eight compared with the values before the breakdown. In other directions, the effect of the breakdowns is not apparent.

The mean and standard deviation values of the QI during the time series are presented in Fig. 9. QI is stable until the MRI scanner gradient breakdown in the month eight. After this breakdown, we see a clearly increased QI values resulting from increased signal outside head area in the image volume.

CNR value mean and standard deviation in the test period are presented in Fig. 10. The value has a decreasing trend throughout the time series. After the gradient breakdown in month eight, the CNR values were substantially decreased.

Fig. 6 MTF10 and MTF50 values measured by using a spherical phantom with variable isotropic acquisition voxel size
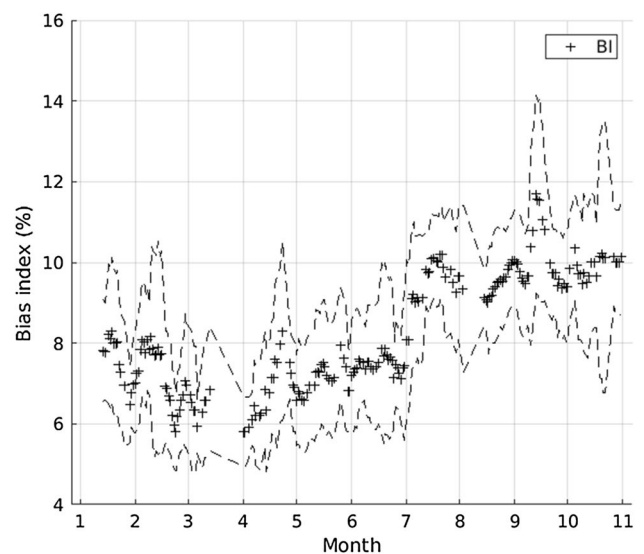

Fig. 7 MTF10 and MTF50 values obtained by using a brain mask filtered with variable sized 3D Gaussian filter
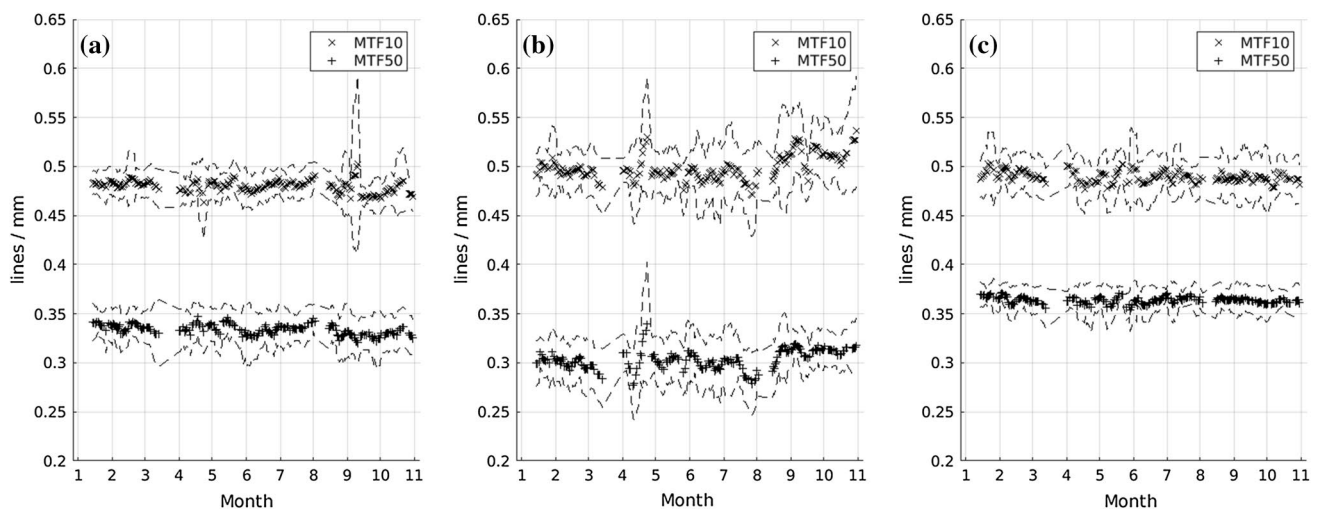
Fig. 8 MTF10 and MTF50 running average and standard deviation (7-day window) in **a** AP, **b** FH and **c** RL directions
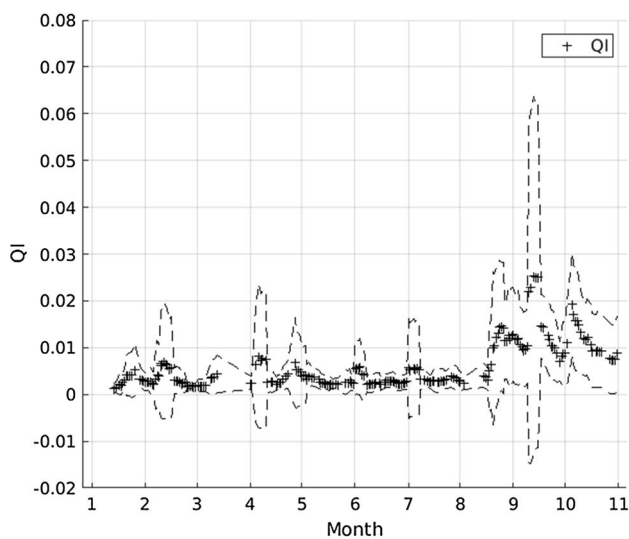
Fig. 9 Quality index running average and standard deviation (7-day window)

Fig. 10 Contrast-to-noise ratio running average and standard deviation (7-day window)

The BI representing the image intensity inhomogeneity is presented in Fig. 11 over the test period. The value increased in month number seven well before the gradient breakdown. At that point, a change of baseline was observed.

Phantom QA result time series for the test period has been presented in Fig. 12. There was no evident effect or trend in the phantom QA results before either of the scanner's breakdowns. A decrease in the variation of the SNR was seen after the second breakdown.

Of the presented QA measures, MTF10, MTF50 and CNR had significantly different results between volumes affected and not affected by patient motion ($p < 0.05$). The effect of patient motion on MTF50 values is presented in Fig. 13.

Discussion

Unexpected changes in image quality are important indicators of an MRI scanner hardware condition. QA measurement can be used to verify nominal the operation of the scanner or in communicating the problems with the manufacturer or service personnel. Aside from detecting errors, it is often important in a patient care setting to know as precisely as possible the date and time when the scanner was
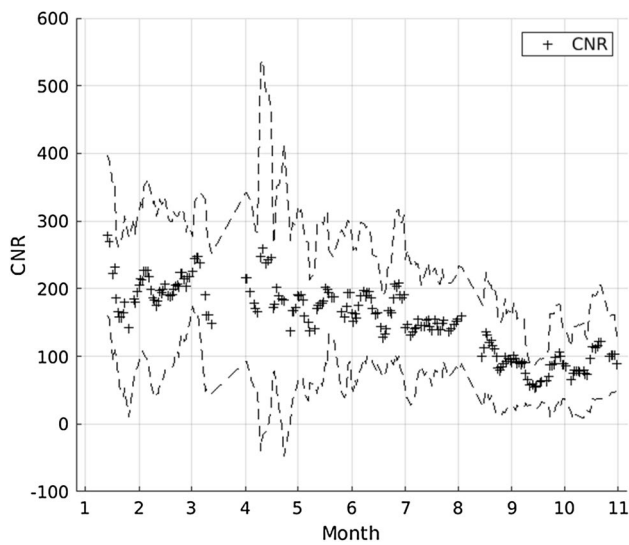
**Fig. 11** Bias index running average and standard deviation (7-day window)

verifiably working properly. It is, however, often difficult to determine if the clinical image quality has degraded during the scanner's lifetime. Several informative parameters can be obtained using standardized QA phantoms and the results compared with those of the acceptance testing or previous quality control measurements. However, the results may not be available, or they may not fully represent the image quality produced by other imaging sequence types or clinical situation. With the presented methods, image quality assessment can be made directly from the clinical image data and results compared with any corresponding study retrospectively. The chosen methods were aimed to be robust, easy to interpret and reflect changes in MRI hardware performance.

The resolution measurement methodology is relying heavily on data averaging to prevent the effect of local anomalies. However, the vast amount of data points in a 3D image is enabling the method of providing information on the actual directional clinical image resolution achieved with the scanner. The accuracy of the resolution assessment was verified with the simulated test models and phantom imaging. High correspondence to idealized expected values was achieved for the degraded brain mask (i.e. simulated) images. Also, a linear relationship between set and measured resolution was observed in the phantom acquisitions verifying the measurement's feasibility for QA purposes.

The MTF10 and MTF50 values are presented here as indicators of MRI scanner resolution. Both values demonstrated a clear response to the changes in scanner hardware. The system breakdowns were seen as a drop of values in FH direction just before the malfunction. The gradient related problems should introduce ambiguity to spatial frequency components and thus the sensitivity to gradient related breakdown is expected. However, the correlation between the RF-system breakdown to changes in MTF values is not apparent. MTF10 values are generally more sensitive to image artefacts with high frequency components. This is seen in the AP direction, where there is a period of increased

Fig. 12 Phantom QA results for the test period. **a** SNR, **b** image intensity uniformity based on the methods presented by NEMA and IEC, **c** image ghosting and **d** phantom diameter in horizontal and vertical direction
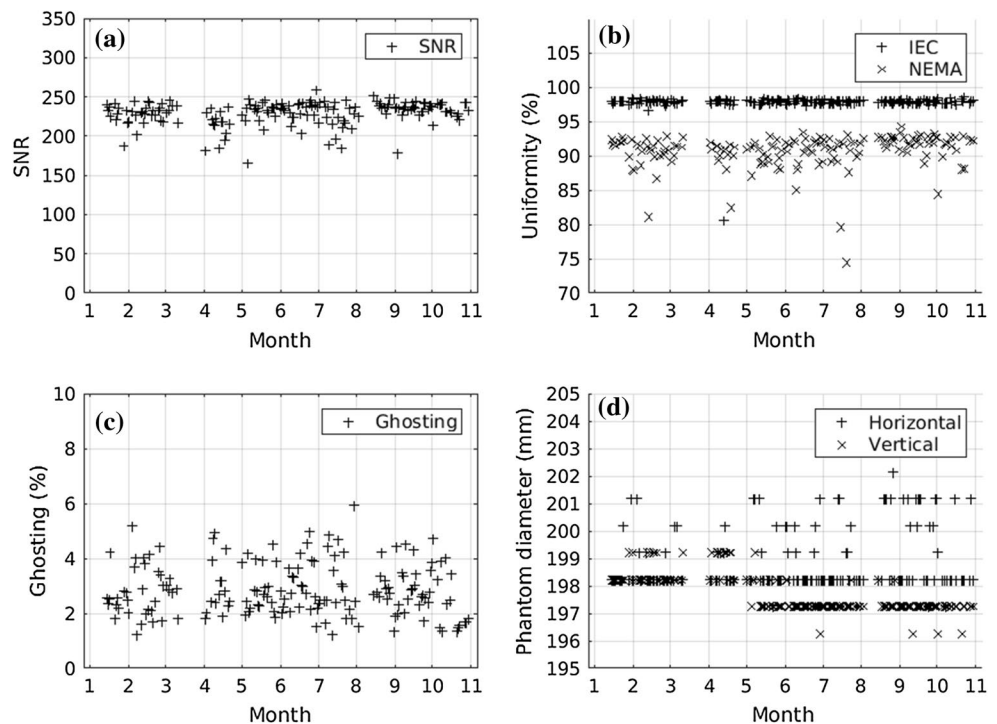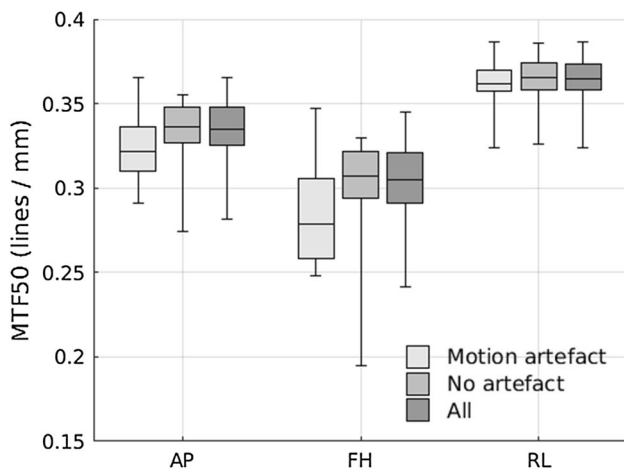
Fig. 13 Median, interquartile range and from the 2nd to 98th percentile range of MTF50 values in three orthogonal directions for studies with motion artefact, without motion artefact and all patients

MTF10 values and standard deviation in month nine that is absent from MTF50 values. A similar period is seen in the RL direction in month four where equal effect is seen in both MTF10 and MTF50 values.

The measurement of QI is based on a principle that all signal outside the actual anatomical volume is resulting from anomalies in the scanning process. Thus, QI is likely to be sensitive to any problems with scanner hardware or patient co-operation. Of the presented quality parameters, QI showed the strongest response on the scanner operation after the second breakdown. QI is not a specific QA measure since there are multiple hardware error mechanisms influencing the value including gradient system instability, the mechanical vibration of a scanner, eddy currents and RF interference. Although no significant correlation between increased QI values and the movement of the patient was found, it is likely that especially major patient movement is provoking increased QI values. This is likely the reason behind small peaks seen in the QI time series.

CNR measurement is depending on the variation of the background noise and contrast between the WM and GM. In the time series we did see a small decreasing trend in the CNR values and a clear drop after the scanner gradient breakdown. The drop after the breakdown is likely caused by increased signal or noise outside the anatomical volume seen also as increased QI values. Likely, the variations in the CNR value are indicating changes in the scanner's RF-system: increased noise or changes in achieved flip angles. Furthermore, patient motion blurs the image and decreases contrast. This is presumably behind the relatively high standard deviation of the CNR values in the time series. The standard deviation could be decreased if only the head volumes without motion would be included.

The BI should be sensitive to inhomogeneities in RF transmit or receive field affecting the RF excitation flip angle in the target volume. These may be induced by problems in RF instrumentation. Additionally, the RF transmission field shape is affected by the anatomical shape of the patient, which may induce strong inherent variation in the measure. Also, metal implants in the patient are causing strong disturbance to the field. A substantial increase in BI value is seen in results during the test period but it is unclear if it has a direct connection to either of the occurred breakdowns. Generally, RF field inhomogeneity effects are effectively corrected by scanner's image intensity normalization algorithms which may decrease methods sensitivity to scanner performance changes. Accessing the normalization algorithm parameters could yield interesting performance information.

The variation of all measurements is highly depending on the stability of image volume segmentation. For the image segmentation, the method is relying on the SPM package which has been utilized and tested comprehensively in multiple studies [31, 32]. If the clinical volume includes severe pathologies, the segmentation algorithm may fail consequently affecting measurement results. Thus, volumes with atypical segmentation results should be removed from the analysis. A simple rule based on GM and WM volume ratio calculated from the segmentation result was used in this study. A more sophisticated set of rules would likely result in decreased standard deviation, but at the same time, limit the amount of available data. The optimization of the criteria for head volume inclusion is should be performed in future.

Additionally, the tetrahedral grid placed on a desired interface to track the perpendicular direction to the surface has to adapt to actual acquisition resolution. The grid has to be dense enough to track the topography of the surface without reacting to voxel-size features. Multiple grid parameter values can be tested empirically to find a stable area where small changes to grid settings have minimal effect on the results.

In addition to physiological error sources, scanner hardware induces inherent variation to the results. For example, the main magnetic field and RF chain can demonstrate temporal fluctuation. One reason to use running average over several days is to mitigate these effects along with interpatient differences.

The presented methods could potentially be utilized as absolute measures to compare the performance of scanners from different vendors and sequence types. The sequence should produce reasonable contrast regions and boundaries, which can be segmented reliably. Other anatomical regions instead of head could also be considered. The effect of changes (e.g. protocol optimization) in sequence parameters on image quality could be studied quantitatively. It is also possible to produce online clinical image quality assurance

tools that are automatically detecting the abnormal operation of the scanner and possibly enable the pre-emptive maintenance of the scanner before noticeable effect on clinical image quality.

Similarly, application of on-line QA enables the detection of patient related issues, e.g. movement. Optimally, this type of system could give a prompt suggestion to rescan before the examination is over. The online detection of a patient motion artefact could significantly affect patient care and costs [33]. The effect of patient motion on the presented QA measures was studied by labeling part of the image volumes as including or not including patient motion related artefact. The motion artefact had significant effect on MTF10, MTF50 and CNR values. In the 3D FLAIR images, the patient motion was found to produce blurring rather than traditional ghosting in the phase-encoding direction. Finding the optimal threshold values for patient motion detection requires further research.

Results obtained from clinical volumes were compared with phantom QA results. There were no clear trends in any of the phantom QA parameters nor visible effects before either of the scanner breakdowns. It is possible that the standard SE sequence is not as demanding in terms of hardware as the 3D FLAIR sequence. Thus, the sensitivity of the phantom QA measurements may not be sufficient to detect effects seen in clinical image QA. Nevertheless, phantom QA has many advantages. For example, it is difficult to evaluate scanner's geometric distortions from patient images.

Recently, machine learning methods have been applied to quality control purposes [34–38]. In future, novel automated approaches may open interesting possibilities in detecting and labeling scanner specific image artefacts. All in all, well-defined, robust, specific and quantitative methods are needed for general QA and sequence optimization purposes, regardless whether they are machine learning or more traditional image analysis in nature.

In this study, four methods for quantifying the image quality of clinical 3D FLAIR acquisitions were presented and applied to a large patient cohort. These can be used in QA to monitor the long-term image quality of an MRI scanner and potentially detect malfunctions before complete hardware failures. The methods can be utilized to measure the effect of changes in sequence parameters or assess the absolute quality of a single patient study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ron P, Jerry A, Geoffrey C, Michael D, Edward HR, Carl K, Jeff M, Moriel NA, Joe O, Donna R (2015) 2015 American College of Radiology MRI Quality Control Manual
2. Fransson A (2000) IPEM Publication, Report No. 80—quality control in magnetic resonance imaging, Lerski R, de Wilde J, Boyce D, Ridgway J, Institute of Physics and Engineering in Medicine, UK, 1999. ISBN 0-904181 901. 36(1). https://doi.org/10.1016/S0720-048X(99)00162-X
3. International Engineering Consortium (2007) IEC 62464-1. Magnetic resonance equipment for medical imaging—part 1: determination of essential image quality parameters
4. National Electrical Manufacturers Association (2008) NEMA standards publication MS 1-2008 determination of signal-to-noise ratio (SNR) in diagnostic magnetic resonance imaging
5. National Electrical Manufacturers Association (2008) NEMA standards publication MS 3-2008 determination of image uniformity in diagnostic magnetic resonance images
6. Lummel N, Schoepf V, Burke M, Brueckmann H, Linn J (2011) 3D fluid-attenuated inversion recovery imaging: reduced CSF artifacts and enhanced sensitivity and specificity for subarachnoid hemorrhage. Am J Neuroradiol 32(11):2054–2060
7. Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 22:5–55
8. Cummins RA, Gullone E (2000) Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. In: Proceedings of the Second International Conference on Quality of Life in Cities, Singapore, pp 74–93
9. Magnotta VA, Friedman L, Birn F (2006) Measurement of signal-to-noise and contrast-to-noise in the fBIRN multicenter imaging study. J Digit Imaging 19(2):140–147
10. Wang W, Hu P, Meyer CH (2007) Estimating the spatial resolution of in vivo magnetic resonance images using radiofrequency tagging pulses. Magn Reson Med 58(1):190–199
11. Mortamet B, Bernstein MA, Jack CR, Gunter JL, Ward C, Britson PJ, Meuli R, Thiran J, Krueger G (2009) Automatic quality assessment in structural brain magnetic resonance imaging. Magn Reson Med 62(2):365–372
12. Borri M, Scurr ED, Richardson C, Usher M, Leach MO, Schmidt MA (2016) A novel approach to evaluate spatial resolution of MRI clinical images for optimization and standardization of breast screening protocols. Med Phys 43(12):6354–6363
13. Osadebey M, Pedersen M, Arnold D, Wendel-Mitoraj K, Alzheimer's Disease Neuroimaging Initiative (2017) The spatial statistics of structural magnetic resonance images: application to post-acquisition quality assessment of brain MRI images. Imaging Sci J 65(8):468–483
14. Jang J, Bang K, Jang H, Hwang D (2018) Quality evaluation of no-reference MR images using multidirectional filters and image statistics. Magn Reson Med 80:914–924
15. Gedamu EL, Collins D, Arnold DL (2008) Automated quality control of brain MR images. J Magn Reson Imaging 28(2):308–319
16. Kruggel F, Turner J, Muftuler LT, Initiative Alzheimer's Disease Neuroimaging (2010) Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49(3):2123–2133

17. Gedamu E (2011) Guidelines for developing automated quality control procedures for brain magnetic resonance images acquired in multi-centre clinical trials. In: Ivanov O (ed) Applications and experiences of quality control. IntechOpen, London, pp 135–158

18. Gedamu EL, Gedamu A (2012) Subject movement during multislice interleaved MR acquisitions: prevalence and potential effect on MRI-derived brain pathology measurements and multicenter clinical trials of therapeutics for multiple sclerosis. J Magn Reson Imaging 36(2):332–343

19. Backhausen LL, Herting MM, Buse J, Roessner V, Smolka MN, Vetter NC (2016) Quality control of structural MRI images applied using FreeSurfer—a hands-on workflow to rate motion artifacts. Front Neurosci 10:558

20. White T, Jansen PR, Muetzel RL, Sudre G, El Marroun H, Tiemeier H, Qiu A, Shaw P, Michael AM, Verhulst FC (2018) Automated quality assessment of structural magnetic resonance images in children: comparison with visual inspection and surface-based reconstruction. Hum Brain Mapp 39:1218–1231

21. Noguchi K, Ogawa T, Inugami A, Toyoshima H, Sugawara S, Hatazawa J, Fujita H, Shimosegawa E, Kanno I, Okudera T (1995) Acute subarachnoid hemorrhage: mR imaging with fluid-attenuated inversion recovery pulse sequences. Radiology 196(3):773–777

22. Bakshi R, Kamran S, Kinkel PR, Bates VE, Mechtler LL, Janardhan V, Belani SL, Kinkel WR (1999) Fluid-attenuated inversion-recovery MR imaging in acute and subacute cerebral intraventricular hemorrhage. Am J Neuroradiol 20(4):629–636

23. Hennig J, Weigel M, Scheffler K (2003) Multiecho sequences with variable refocusing flip angles: optimization of signal behavior using smooth transitions between pseudo steady states (TRAPS). Magn Reson Med 49(3):527–535

24. Busse RF, Hariharan H, Vu A, Brittain JH (2006) Fast spin echo sequences with very long echo trains: design of variable refocusing flip angle schedules and generation of clinical T2 contrast. Magn Reson Med 55(5):1030–1037

25. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66

26. Van Den Boomgaard R, Van Balen R (1992) Methods for fast morphological image transforms using bitmapped binary images. CVGIP Graph Models Image Process 54(3):252–258

27. Steckner MC (1994) Computing the modulation transfer function of magnetic resonance imagers. Dissertation, The University of Western Ontario

28. Fang Q, Boas DA (2009) Tetrahedral mesh generation from volumetric binary and grayscale images. In: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, pp 1142–1145

29. Canny J (1987) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intel 8(6):679–698

30. Peltonen JI, Mäkelä T, Sofiev A, Salli E (2017) An automatic image processing workflow for daily magnetic resonance imaging quality assurance. J Digit Imaging 30(2):163–171

31. Chard DT, Parker GJ, Griffin C, Thompson AJ, Miller DH (2002) The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology. J Magn Reson Imaging 15(3):259–267

32. Kazemi K, Noorizadeh N (2014) Quantitative comparison of SPM, FSL, and brainsuite for brain MR image segmentation. J Biomed Phys Eng 4(1):13–26

33. Andre JB, Bresnahan BW, Mossa-Basha M, Hoff MN, Smith CP, Anzai Y, Cohen WA (2015) Toward quantifying the prevalence, severity, and cost associated with patient motion during clinical MR examinations. J Am Coll Radiol 12(7):689–695

34. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JL, Griffanti L, Douaud G, Sotiropoulos SN, Jbabdi S, Hernandez-Fernandez M, Vallee E (2018) Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. Neuroimage 166:400–424

35. Esteban O, Birman D, Schaer M, Koyejo O, Poldrack R, Gorgolewski K (2017) MRIQC: predicting quality in manual MRI assessment protocols using no-reference image quality measures. BioRxiv 27:2017

36. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ (2017) MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One 12(9):e0184661

37. Pizarro RA, Cheng X, Barnett A, Lemaitre H, Verchinski BA, Goldman AL, Xiao E, Luo Q, Berman KF, Callicott JH (2016) Automated quality assessment of structural magnetic resonance brain images based on a supervised machine learning algorithm. Front Neuroinform 10:52

38. Küstner T, Liebgott A, Mauch L, Martirosian P, Bamberg F, Nikolaou K, Yang B, Schick F, Gatidis S (2017) Automated reference-free detection of motion artifacts in magnetic resonance images. Magn Reson Mater Phy 31(2):243–256