



Published in final edited form as:

*Magn Reson Imaging*. 2017 January ; 35: 81–90. doi:10.1016/j.mri.2016.08.022.

## Quantitative Quality Assurance in a Multicenter HARDI Clinical Trial at 3T

Xiaopeng Zhou<sup>a</sup>, Ken E. Sakaie<sup>a</sup>, Josef P. Debbins<sup>b</sup>, John E. Kirsch<sup>c</sup>, Curtis Tatsuoka<sup>d</sup>, Robert J. Fox<sup>e</sup>, and Mark J. Lowe<sup>a</sup>

<sup>a</sup>Imaging Institute, Cleveland Clinic, Cleveland, OH 44195, USA

<sup>b</sup>Barrow Neurological Institute, St. Joseph's Hospital and Medical Center, Phoenix, AZ 85013, USA

<sup>c</sup>Siemens Medical Solutions, Cary, NC 27511, USA

<sup>d</sup>Department of Neurology, Case Western Reserve University, Cleveland, OH 44106, USA

<sup>e</sup>Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, USA

### Abstract

A phantom-based quality assurance (QA) protocol was developed for a multicenter clinical trial including high angular resolution diffusion imaging (HARDI). A total of 27 3T MR scanners from 2 major manufacturers, GE (Discovery and Signa scanners) and Siemens (Trio and Skyra scanners), were included in this trial. With this protocol, agar phantoms doped to mimic relaxation properties of brain tissue are scanned on a monthly basis, and quantitative procedures are used to detect spiking and to evaluate eddy current and Nyquist ghosting artifacts. In this study, simulations were used to determine alarm thresholds for minimal acceptable signal-to-noise ratio (SNR). Our results showed that spiking artifact was the most frequently observed type of artifact. Overall, Trio scanners exhibited less eddy current distortion than GE scanners, which in turn showed less distortion than Skyra scanners. This difference was mainly caused by the different sequences used on these scanners. The SNR for phantom scans was closely correlated with the SNR from volunteers. Nearly all of the phantom measurements with artifact-free images were above the alarm threshold, suggesting that the scanners are stable longitudinally. Software upgrades and hardware replacement sometimes affected SNR substantially but sometimes did not. In light of these results, it is important to monitor longitudinal SNR with phantom QA to help interpret potential effects on in vivo measurements. Our phantom QA procedure for HARDI scans was successful in tracking scanner performance and detecting unwanted artifacts.

---

Address correspondence to: Mark Lowe, PhD, Imaging Institute, Cleveland Clinic, 9500 Euclid Ave, U15, Cleveland, OH 44195, USA, lowem1@ccf.org, Phone: 216 445 2661, Fax: 216 636 2397.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Keywords

multicenter; QA; HARDI; SNR; phantom

---

## 1. Introduction

Multiple sclerosis (MS) is a chronic disease characterized by inflammatory demyelination of the central nervous system. Ibutilast is an agent that has shown potential neuroprotective efficacy in MS [1]. The Secondary and Primary pRogressive Ibutilast Neuronext Trial (SPRINT-MS), which is being conducted at 28 clinical sites to ensure high patient enrollment and diversity in the study population, is using the NeuroNEXT Network ([www.neuronext.org](http://www.neuronext.org)), a National Institutes of Health-sponsored framework designed to facilitate clinical trials of treatment for neurological diseases. If this trial demonstrates that Ibutilast is effective in slowing the progression of atrophy or other advanced imaging measures of neurodegeneration, it would represent a significant step forward in the development of therapy for progressive MS.

SPRINT-MS is using advanced magnetic resonance imaging (MRI) to characterize brain tissue integrity in patients with progressive MS and to correlate imaging measures with clinical activity. The primary outcome of this trial is whole-brain atrophy. A secondary outcome of this trial is change in diffusivity measurements as measured by high angular resolution diffusion imaging (HARDI) [2]. The quantitative nature of HARDI makes it attractive for multicenter clinical trials, as this technique can characterize brain tissue integrity with high granularity and may be useful for measuring the benefit of putative neuroprotective therapies [3]. In the SPRINT-MS trial, change in transverse diffusivity (TD) along the pyramidal tracts is being evaluated as a biomarker for the efficacy of Ibutilast treatment.

Variability among scanners may cancel the benefit of using multiple centers to assess new treatments. These scanner variabilities may be attributed to differences among scanner hardware and software. Previous work demonstrated that comparable fractional anisotropy and diffusivity values could be obtained from 5 different 3T MR scanners and platforms, even with a software upgrade [4]. However, other studies have found that imaging measurements may differ when various models of scanners or scanners from various manufacturers are used [5] or even when the same scanner model is used [6, 7]. Therefore, differences among scanners, including differences caused by repairs or upgrades of software or hardware, must be quantified in longitudinal studies. Standardizing protocols, training technologists in scanning techniques and periodically performing quality assurance (QA) for each scanner are essential when attempting to minimize differences among study centers.

A number of artifacts may degrade HARDI image quality and lead to inaccurate quantitative measurements. A minimum signal-to-noise ratio (SNR) is required in diffusion-weighted (DW) imaging to prevent systematic bias in diffusivity values [8]. HARDI sequences require many gradient directions with large durations and amplitudes [9] which, together with the fast switching of gradients in the echo planar imaging (EPI) readout, can provoke spiking artifact more frequently than in conventional imaging [10]. The combination of large

diffusion-weighting gradients and low bandwidth (BW) of the EPI readout risks severe geometric distortion due to eddy currents [11]. Because EPI readout uses a zigzag trajectory through k-space, Nyquist ghosting artifacts can also occur on reconstructed images [12].

Although a standard functional MRI (fMRI) QA protocol and criteria have been published [13], there is no universally accepted standard QA procedure for HARDI. Previous multicenter diffusion imaging studies have been limited to scanners from the same manufacturer or to only a few sites [14] or fewer gradient directions [5, 15].

In this study, we developed a phantom-based QA protocol for a multicenter HARDI clinical trial using 3T MR scanners from 2 major manufacturers at 27 imaging sites. We used quantitative procedures to detect spiking and to evaluate eddy current and Nyquist ghosting artifacts. Simulations were used to determine alarm thresholds for minimal acceptable SNR.

## 2. Materials and Methods

### 2.1. MR Scanners

At the beginning of this study, a detailed survey was sent to each site in the NeuroNEXT Network to gather information about the essential features of readily accessible scanners (manufacturer, model, field strength, head coil, software, ability to perform HARDI). A total of 27 MR scanners (11 Siemens TIM Trio, 6 Siemens Skyra, 1 GE Signa EXCITE, 7 GE Signa HDxt, 1 GE DISCOVERY MR750, and 1 GE DISCOVERY MR750W) were approved for inclusion in the study (Table 1). The scanners were manufactured by Siemens (Erlangen, Germany) and GE (Waukesha, Wisconsin, USA). The scanners had various software levels (Siemens: VB17 and VD13; GE: 12x, 15x, 16x, 23x, and 24x).

Head coil choice was constrained to limit variability due to sensitivity profiles of the coils. Standard 20-, 12-, and 8-channel coils were required on Siemens Skyra, Siemens Trio, and GE scanners, respectively. However, 1 GE Discovery MR750 site (scanner #1) was limited to a 16-channel (HNS HEAD) coil.

### 2.2 Site Visits

Two MRI physicists visited each site to train the technologists on phantom QA procedures (described below), which were designed to limit variability in scanning the phantom. A healthy control qualifying (HCQ) scan was also acquired at the time of the visit. The protocols can be found in the supplemental material. Both the HCQ and initial phantom qualifying scan had to be approved before the site was allowed to recruit patients.

### 2.3. Phantom Scans

The BIRN phantom was chosen as the standard phantom for this trial because it is readily available and because its properties match those of brain tissue [13]. Five imaging sites (sites #11, #18, #21, #23, and #27 in Table 1) used their own BIRN phantoms for this trial. For the remaining 22 sites, phantoms were bought and shipped to each site before the site visit.

A procedure was developed to ensure consistent placement of the phantom. For Siemens scanners, standard phantom holders (part #10606579, no. 29 and part #7581700, no. 14)

were used with dedicated adapter rings constructed with a 3D printer (Fortus 250mc, StrataSys, Eden Prairie, Minnesota, USA) (Fig. 1). The GE scanners have a standard ring holder on the head coil to hold the phantom in position (Fig. 1). To ensure consistent positioning, lines were drawn along the laser landmarks on each phantom during the site visit. The technologists were instructed to use these lines to ensure reproducible positioning.

#### 2.4. HARDI Sequences

Previous research has demonstrated that matching HARDI parameters leads to high concordance in DTI parameters despite hardware differences [4]. Comparable HARDI protocols were generated for each system (flip angle =  $90^\circ$ , isotropic resolution = 2.5 mm, field of view (FOV) =  $255 \text{ mm} \times 255 \text{ mm} \times 150 \text{ mm}$ , phase partial fourier factor = 6/8, no parallel imaging). The HARDI sequence included 8  $b = 0$  volumes and 64 noncollinear DW gradients with b-value of  $700 \text{ s/mm}^2$ , matching the profile used in previous research [16]. One Signa Excite scanner (scanner #3) could acquire only 55 DW volumes. Trio scanners used twice-refocused spin-echo diffusion weighting [17], whereas Skyra and GE scanners used Stejskal-Tanner diffusion weighting [18]. The twice-refocused spin-echo sequence limits eddy current distortions. The Stejskal-Tanner sequence allows a shorter echo time and hence higher SNR but is not standard on Trio scanners. Detailed protocols can be found in the supplemental material.

Most MR units have heavy demands on scan time, so limiting the frequency and duration of QA testing helps to ensure adherence. Phantom QA scans are performed monthly. The scan protocol matches that of in vivo scans but uses fewer slices and a shorter TR to limit the scan time. For phantom QA scans, the scan duration was 6 minutes to acquire 36, 39, and 21 slices for Trio, Skyra, and GE scanners, respectively.

#### 2.5. Volunteer Scans

Volunteer scans were approved by a central Institutional Review Board (IRB), and informed consent was obtained from each volunteer. Volunteers aged 22 to 56 years (16 females, 11 males) were scanned at each center so that technologists could be trained on the procedures for subject positioning. However, volunteers could not be scanned at several sites at the time of training because of a lack of a DTI license on the scanner (#17), pending IRB approval (#21), or a scanner undergoing repairs (#25). At all other sites, whole-brain T1-weighted images were acquired with the following parameters: 192 axial slices, slice thickness = 1 mm, FOV =  $256 \times 256 \text{ mm}^2$ , repetition time (TR)/echo time (TE) = 20/6 ms, flip angle =  $27^\circ$ , and BW = 160 Hz/pixel (Siemens) or 22.73 kHz (GE). For the T1-weighted sequence, technologists were instructed to use an oblique sequence when necessary. However, a key point of the HARDI sequence was to prescribe straight axial slices. For volunteer scans, HARDI required approximately 10 minutes for whole-brain coverage. The HCQ scans were repeated within 24 hours in the same participants to assess measurement reproducibility.

#### 2.6. Comparison of HARDI Sequences

The choice of HARDI sequence on each type of scanner was based on the availability of sequence, optimization of SNR and matching across different platforms. However, there were still discrepancies among HARDI sequences that may have affected our results. To test

how much the selected protocol and parameters affected comparisons among scanners, additional phantom and volunteer scans were conducted on a Skyra scanner at one institute, which was not included in this trial. For the phantom scans, a range of parameter values used across imaging platforms were examined. The influence on SNR and eddy current distortion of pulse sequence (twice-refocused spin-echo and Stejskal-Tanner), TE (69–93 ms), TR (4000/4500 ms, corresponding to GE/Siemens phantom QA scans and BW (1885, 1960, and 2040 Hz/pixel) were examined. For the volunteer scans, three different parameter configurations, corresponding to those of in vivo scans across imaging platforms were used (TR/TE: 7400 ms/80 ms, corresponding to Trio, TR/TE: 7000 ms/69 ms, corresponding to Skyra, TR/TE: 8500 ms/93 ms, corresponding to GE).

## 2.7. Data Analysis

DICOM headers for all of the scans were checked for protocol compliance. Analysis of Functional NeuroImages (AFNI) [19] was used to qualitatively check for artifacts and correct phantom positioning. Quantitative analysis was used to assess spiking [10], eddy current [20], Nyquist ghosting [21] and SNR. Quantitative analyses were processed using software written in-house with a commercial package (MATLAB R2012a, The MathWorks Inc., Natick, MA). All in vivo HARDI images were eddy current and motion corrected using Tortoise [22].

## 2.8. Spiking Artifact

Our QA protocol incorporated an automated method to objectively detect and localize spiking artifact on HARDI data. Spiking artifact appears as ripples on MR images and can arise from a number of sources such as electrical arcing and spurious static discharge. Severity ranges from obvious to negligible. Severe spiking artifact can be identified with a visual check (Fig. 2). Even mild spiking can lead to inaccuracies in quantitative metrics such as diffusivity, and negligible or low-level spiking artifact may indicate problems that can progressively deteriorate. Therefore, automatic objective spiking detection is important to detect low-level spiking artifact. For this study, we used Outlier Detection De-spiking (ODD) [10].

## 2.9. Eddy Current Correction

An iterative cross-correlation (ICC) postprocessing algorithm [11, 20] was used to quantify and correct eddy current-induced artifacts in terms of scale, shear, and translation in DW images compared with non-DW images. Eddy currents can generate a dynamic magnetic field error that combines with the gradient prescribed by the specific sequence, leading to distortion in DW images. The ICC method can estimate distortions in DW images by cross-correlating the DW images with non-DW images and can correct the distortion using the estimated distortion parameters.

## 2.10. SNR Measurement

We measured the  $\text{SNR}_{b_0}$  (the SNR among images without diffusion weighting) within a square region of interest (ROI) of  $4000 \text{ mm}^3$  placed automatically by the software at the magnetic isocenter of the central slice. This was possible because the technologists were

trained to prescribe the QA scan such that the magnetic isocenter was within the central slice of the phantom. SNR was calculated as follows. For each pixel in the ROI, there were 8 measurements without diffusion weighting ( $b = 0$ ). The mean and standard deviation of these 8 measurements were calculated for each pixel. The mean was averaged across pixels to estimate the signal, and the standard deviation was averaged across pixels to estimate the noise.  $\text{SNR}_{b_0}$  is the ratio of these estimates of signal and noise.

The SNR for DW images ( $\text{SNR}_{\text{DW}}$ ) was calculated to determine the risk of noise floor bias in diffusivity, which can occur for  $\text{SNR}_{\text{DW}} < 2$  [16, 23, 24]. As noise floor rectification biases low  $\text{SNR}_{\text{DW}}$  measurements, it is unreliable to measure SNR directly from DW images.  $\text{SNR}_{\text{DW}}$  was estimated as follows:

$$\text{SNR}_{\text{DW}} = \text{SNR}_{b_0} \exp(-bD), \quad (1)$$

in which  $b = 700 \text{ s/mm}^2$  and  $D$  is set to a worst-case value of  $1.5 \times 10^{-3} \text{ mm}^2/\text{s}$ , the value along the principal eigenvector of highly organized white matter.  $\text{SNR}_{\text{DW}} > 2$  was deemed acceptable. In vivo SNR was calculated from the 8  $b = 0$  images, similarly to how SNR was calculated for the phantom for all voxels within a white matter mask. The mask was constructed by segmenting T1-weighted images using the FAST algorithm from the FSL library [25, 26].

### 2.11. Percent Signal Ghosting

Percent signal ghosting (PSG) was used to quantify and monitor Nyquist ghosting [21]. Two rectangular ROIs were drawn on a maximum intensity projection to estimate background noise and ghost artifact. The mean signal was calculated from a square ROI at the isocenter described above, and PSG was calculated as follows:

$$\text{PSG} = 100\% \times |(\text{ghost signal} - \text{background signal}) / (2 \times \text{mean signal})| \quad (2)$$

A PSG value  $< 3\%$  was deemed acceptable [21].

### 2.12. SNR Alarm Threshold

We developed a simulation framework to determine an alarm threshold for SNR. SPRINT-MS is a longitudinal study with one objective being measurement of change over time in diffusion tensor imaging (DTI) measures of tissue microstructure in each subject on each scanner. One DTI measure, TD (also known as radial diffusivity), correlates consistently with a hallmark of MS, demyelination [27]. We expect TD to increase over time if demyelination occurs and to be stable over time if myelin status is stable. For the SPRINT-MS study, it is hypothesized that TD will increase over time in the untreated placebo group and remain stable over time in the treatment group. A key metric of scanner performance in this study is sensitivity to a 10% difference in TD (henceforth referred to as “sensitivity”). This magnitude difference is typical of differences between values in the white matter of healthy controls and MS patients [28] and is therefore assumed to be appropriate when assessing longitudinal changes. The sensitivity should remain stable over time. A decline in

sensitivity over time may prevent detection of a change in TD. However, it is difficult to measure sensitivity to change directly. Simulations can provide a connection between SNR, which is straightforward to measure, and sensitivity. Details of the simulation framework are provided in the supplementary material. For each scanner, the average SNR of the first 3 phantom measurements was used to establish a baseline SNR. The simulation framework relates this SNR to a baseline value of sensitivity. Subsequent monthly measurements of SNR are related to subsequent values of sensitivity. If the sensitivity drops by 10% from the baseline sensitivity for a given site, the site repeats the measurement for verification. If the sensitivity is still low, a field service engineer visits the site to evaluate the scanner.

### 3. Results

#### 3.1. Incidence of Repeated HCQ and Phantom QA Scans

A total of 4 centers had to repeat qualifying phantom QA scans because of noncompliance with the acquisition protocol or spiking artifact, and 7 centers had to repeat HCQ scans because of incorrect prescription of the FOV, spiking artifact, fat ring artifact or motion artifact. Overall, 9 centers had problems with qualifying scans (Table 2). There were 50 repeated monthly QA scans out of 361 QA scans over 1.5 years because of artifacts or operator error. A common operator error was incorrect positioning of the phantom. On Siemens scanners, 7 of the 8 non-DW images are acquired as separate scans. Inconsistent positioning among DW and non-DW images occurred.

#### 3.2. Spiking, Nyquist Ghosting Artifact, and Eddy Current Distortion on Phantom Scans

Three instances of subtle spiking artifact were observed in the qualifying scans from 3 scanners (#24, #25, #26). These scanners were subsequently fixed. Subsequent monthly QA scans detected spiking artifact on 2 other scanners (#19, #21). Imaging was suspended at these sites until repairs were completed.

One Trio scanner (#22) demonstrated minor Nyquist ghosting artifact intermittently on DW volumes (Fig. 3). PSG values in the fourth, sixth, seventh and eighth months were abnormal, although only the PSG value in the fourth month (Fig. 3, Box 4) was higher than the threshold. Another Trio scanner (#25) showed severe ghosting artifact on all volumes at 3 time points (Fig. 3, Box 11–13). PSG values were much lower (Fig. 3, Box 9,10,14–18) after the scanners were repaired and were similar to PSG values from scanners of the same type (Fig. 3, Box 19–23). A surprisingly common cause of excessive ghosting was unbalanced voltage input to the gradient amplifiers. Such problems were readily addressed by the field service engineer.

Overall, Trio scanners exhibited less eddy current distortion than GE scanners, which in turn showed less distortion than Skyra scanners (Fig. 4). Skyra scanners generally showed more eddy current distortion than other types of scanners, but the degree of distortion was stable (Fig. 5a). After the standard eddy current correction procedure was applied to a Trio scanner (#25) (Fig. 5b), the QA data showed less eddy current distortion than was seen in the qualifying phantom scan. One Discovery MR750W scanner (#2) showed more serious image distortions than other GE scanners in the qualifying scan (Fig. 5c). This scanner used

a different head coil for the first 2 QA scans (Head 24). After the coil was changed to the standard 8-channel receive coil (8HRBRAIN), the image distortions were reduced and distortion parameters were stable for 15 months.

### 3.3. Comparison of HARDI Sequences on a Single Scanner

When HARDI sequences were compared on a Skyra scanner with a phantom, there was less eddy current distortion with the twice-refocused spin-echo sequence than with the Stejskal-Tanner sequence. With the Stejskal-Tanner sequence using both TR 4000 ms and 4500 ms, there was less eddy current distortion when TE was longer, which may explain why less eddy current distortion was seen with GE scanners (TE = 93 ms) than with Skyra scanners (TE = 69 ms). BW had little effect on eddy current distortion. A TR change from 4000 ms (GE) to 4500 ms (Skyra and Trio) did not greatly affect SNR, whereas SNR decreased with increasing TE for both the Stejskal-Tanner sequence and the twice-refocused spin-echo sequence with different TRs. Although there was a slight decrease in SNR when the BW was increased from 1885 Hz/pixel to 1960 Hz/pixel, BW did not affect SNR when the BW was changed from 1960 Hz/pixel to 2040 Hz/pixel. Because of software limitations, we could not increase the BW to 2130 Hz/pixel to match the BW used in the Trio protocol. Results from the volunteer scan matched those from the phantom scan: the twice-refocused spin-echo sequence demonstrated less eddy current distortion, and the SNR was higher with shorter TEs.

### 3.4. SNR Measurements

There was a linear relationship between phantom SNR and in vivo SNR (Fig. 6a). In vivo SNR measurement reproducibility from the volunteer scan-rescan study was calculated with the Bland-Altman method (Fig. 6b). There was high agreement for SNR values on the scan and rescan ( $R^2 = 0.8$ ).

The mean SNR measurements from the qualifying phantom scans varied among scanners (Fig. 7a), which may have been the result of the different protocols used. In general, SNR was similar among GE and Trio scanners but was much higher for Skyra scanners. A Discovery MR750 scanner (#1) and one Signa HDxt scanner (#4) showed higher SNR than other GE scanners. Fig. 7b shows an example using the threshold defined by the simulation framework described in the supplementary material. For one Trio scanner (#17), the SNR in the ninth month was below the alarm threshold. A repeat QA scan confirmed that there were no hardware problems. The SNR alarm threshold from all phantom QA scans was breached 3 times after datasets with artifacts and position errors were excluded. In each case, the phantom was scanned again. In 2 cases, the rescan SNR was above the alarm threshold, obviating the need for further action. In the third case, the site was asked to have a field service engineer examine the system. Subsequent monthly scans demonstrated SNR values above the alarm threshold for all of the sites.

### 3.5. Impact of Software Upgrade and Hardware Replacement on SNR

During this study, there were several instances of software and hardware upgrades and hardware replacement. Table 3 summarizes the effect of these changes on SNR. We expected hardware changes but not software changes to result in large changes in SNR, but this was



not what we observed. Substantial changes (replacement of gradient coil or head coil) sometimes led to only slight changes in SNR, whereas one software upgrade led to a large decrease in SNR.

### 3.6. Longitudinal Stability of Diffusion Metrics and the Effect of Uniformity Within and Among phantoms on Diffusion Metrics

We investigated the quality and longitudinal stability of the entire pipeline of acquisition and reconstruction with respect to diffusion metrics. For each scan, we calculated MD and FA in each voxel. The MD and FA values were then averaged across the ROI used for the SNR measurement. The temporal variation of these diffusion metrics was then calculated by taking the mean and standard deviation across values from monthly scans at each site. The ratio of the temporal standard deviation and mean, the temporal coefficient of variation ( $CV = 100 (SD/mean) \%$ ) was then used to summarize the stability of the diffusion metrics. Values of the temporal mean and standard deviation of MD and FA from measurements taken over 1.5 years are shown in Fig. 8. Among all sites, the temporal coefficient of variation of MD ranged from 0.4% to 5.8%. The temporal CV of FA ranged from 1.8% to 22.7%. To assess the impact of nonuniformity within each phantom, we repeated the evaluation in Fig. 8 at ROIs at two slices adjacent to the central slice. Among all sites, the maximum change of temporal CV values when varying the location of the ROI were 0.2% for MD (a change in CV from 1.0% to 0.8%) and 4.6% for FA (a change in CV from 3.2% to 7.8%). To assess the non-uniformity among the 27 phantoms, we calculated the mean ( $M_{inter}$ ) and standard deviation ( $SD_{inter}$ ) across sites of the temporal means of MD and FA described above. The inter-site coefficient of variation ( $CV_{inter} = 100 (SD_{inter}/M_{inter}) \%$ ) were 6.3% for MD and 29.1% for FA.

## 4. Discussion

We endeavored to establish a practical, easily implemented HARDI QA protocol to evaluate MR hardware, software and operator performance in a longitudinal, multicenter trial. This study demonstrated that our QA protocol can effectively detect scanner malfunctions and operator errors in many situations. The QA measurements can also indicate whether there have been minor or major changes caused by hardware repair or replacement and software and hardware upgrades.

Initial qualifying scans and monthly QA scans can be invaluable in identifying scanner problems and operator errors. The most common operator errors are incorrect slice prescription for in vivo scans and incorrect positioning of the BIRN phantom. Rapid feedback helps resolve these issues. Additionally, when artifact was detected on phantom scans, subject scans were suspended until repairs to the scanner were completed. This prevented the need for repeat subject scans, which are costly and burdensome on patients and imaging centers.

Our results showed that spiking artifact was the most common type of artifact. Trio scanners employ twice-refocused spin-echo diffusion weighting, whereas the other scanners use Stejskal-Tanner diffusion weighting. The twice-refocused spin-echo sequence may be more

demanding on the gradient hardware due to additional gradient lobes, perhaps explaining this higher incidence of spiking.

Our QA with the ODD method [10] is sensitive to subtle spiking artifact. Most cases of spiking artifact found on the monthly QA scans were not severe. However, some of the subtle artifacts were not easily detected with a visual check. Additionally, adjusting the contrast to inspect cases of spiking artifact image by image is time consuming. With the quantitative method, subtle spiking artifact can be detected objectively and quickly.

Nyquist ghosting artifact is another artifact that was seen on images from 2 scanners. One instance of Nyquist ghosting artifact was severe on both DW and non-DW images and could be easily identified with a visual check. The other intermittent Nyquist ghosting artifact was difficult to discern. The use of the PSG parameter allows for easier identification of these artifacts on certain volumes.

We used an ICC-based method to quantify eddy current distortion [20]. This approach is easy to implement and does not require the acquisition of additional images. The derived parameters (shear, scale, and translation) are physically meaningful and are related to gradients in the readout, phase and slice direction, respectively. As we expected, Trio scanners demonstrated the least distortion, as the twice-refocused spin-echo sequence was designed to limit eddy current artifact. Skyra scanners demonstrated the highest level of eddy current distortion, which was also expected because these scanners use Stejskal-Tanner diffusion weighting. However, the GE scanners did not demonstrate the same degree of distortion as the Skyra scanners despite the use of the same type of Stejskal-Tanner diffusion weighting. This difference may be due to the design of the gradient coils or TE. Our comparison scanning performed in a phantom and a volunteer demonstrated that when the Stejskal-Tanner sequence was used, the eddy current distortion was lower with higher TEs. The large difference between the TE used on Skyra scanners (69 ms) and the TE used on GE Signa HDxt scanners (93 ms) may explain this difference in eddy current distortion on the phantom, although this difference was not obvious in the volunteer scans. However, for a longitudinal study, we expect that the stability of distortion parameters over time will be more important than differences among scanners.

In previous work, we showed that there is an inverse relationship between SNR and diffusivity and that the dependency of DTI measures on SNR is different among various scanner types [29]. Therefore, SNR is an important parameter in this QA procedure designed to monitor scanner stability. The volunteer scan-rescan study demonstrated high SNR reproducibility, and there was a strong linear relationship between phantom SNR and in vivo SNR. Skyra scanners showed the highest SNR because they used the shortest TE. Trio and GE scanners demonstrated similar SNRs even though the TE used on Trio scanners was shorter than that used on GE scanners.

We developed alarm thresholds for SNR tailored to one of the outcomes of SPRINT-MS: changes in TD over time along pyramidal tracts. Although we observed considerable variability among scanner platforms in SNR, minimum levels were sufficient to avoid bias in diffusivity due to noise floor rectification in all cases. One threshold was derived from well-

established analyses of Rician noise [16, 24]. The simulation framework translates SNR, which is easily measured, into a quantity directly relevant to this longitudinal study: changes in the sensitivity to changes in TD. Nearly all of the phantom measurements with artifact-free images were above the alarm threshold, suggesting that the scanners are stable longitudinally.

Scanner upgrades are a common source of concern for any longitudinal study. There is currently no completely satisfying way to address these issues. In a previous study, a software upgrade from HDx to HDxt for a GE scanner had a significant effect on DTI measurements [6]. However, a Siemens scanner software upgrade was associated with only a slight change in fractional anisotropy measurements [4]. The phantom scans in our study showed that hardware and software changes sometimes affected SNR substantially but sometimes did not. Discovery MR750 (with HNS head coil and software 23x) and one Signa HDxt with software 15x showed higher SNR than other GE scanners, which may have been caused by these hardware and software differences. We could not establish the root cause for these changes because of limitations on resources. The regularly scheduled phantom scans do indicate when substantial changes in system performance may have occurred. Such information may prove useful to control for variability in SPRINT-MS caused by instrument changes as opposed to biological changes.

The BIRN phantom, although designed to match the imaging characteristics of brain tissue, is not an ideal standard for diffusion MRI. The diffusivity of the doped agar gel ( $1.5\text{--}1.8 \times 10^{-3} \text{ mm}^2/\text{s}$ ) does not overlap the range of values found in brain tissue ( $0.5\text{--}1.4 \times 10^{-3} \text{ mm}^2/\text{s}$ ) [28]. Additionally, uniformity among agar phantoms is more difficult to achieve than among liquid phantoms, and stability over long periods of time can be a concern. We are aware of only one other multi-center study that evaluated the stability of diffusion metrics in phantoms over time. Zhu et al. measured the intra-site variance of diffusion metrics of a single phantom consisting of 3 cyclic alkanes on 3 GE scanners with the same hardware and software configuration over 6 weeks [14]. From their measurements, we estimate values of temporal CV of MD ranging from 8.9% to 14.6%. For isotropic phantoms, FA is expected to be close to zero. Assessing the CV of FA may be unenlightening because small differences in the denominator of the equation for CV will cause large differences in the overall value. We therefore focus on the temporal standard deviation of FA, which ranges from 0.0008 to 0.0079 in our study. The measurements of intra-site variance of FA in the study of Zhu et. al [14] places an upper limit on temporal standard deviation ranging from 0.13 to 0.16. Our measurements of stability reflect time variation of the scanner hardware and of the composition of the agar phantoms. That the values do not exceed those found in the highly controlled study of Zhu et al. suggests that variability due to the agar phantoms themselves is not a major factor.

ROI placement had little effect on the measures of stability of MD and FA, suggesting that non-uniformity of the agar within each phantom was not an important concern. Inter-scanner CV values of 6.3% for MD 29.2% for FA were comparable to those found in an ice-water phantom (3% for MD and 27.8% for FA) [30] suggesting that the composition of the agar was comparable across sites.

The ACR phantom [15] and ice-water phantom [31] have been used in previous multicenter quality control studies for diffusion imaging. However, the diffusivity of the ACR phantom ( $\sim 3 \times 10^{-3} \text{ mm}^2/\text{s}$ ) is even more poorly matched to the diffusivity of tissue than the BIRN phantom. The ice-water phantom matches the tissue values of diffusivity more closely, but this phantom must sit for approximately 30 minutes to achieve thermal equilibrium. Liquid phantoms also need to sit in the scanner for several minutes to prevent flow artifacts. Because of the busy schedule of most clinical imaging centers, it is not practical to use a phantom that requires a long preparation time. Organic diffusion standards [32] are impractical in a multicenter trial because of toxicity and volatility issues. Polyvinylpyrrolidone phantoms are a promising alternative [33], but these phantoms were not available at the initiation of the SPRINT-MS trial.

One artifact that has arisen among in vivo scans that the phantom scans do not address is the fat-ring artifact. This artifact results from either patient motion or inadequacy of the fat saturation pulse. As the phantoms do not contain fat, we cannot use the phantom scans to distinguish between these causes of the artifact. This is an unfortunate case in which phantom scans cannot prevent rescans of human patients.

There are similarities between our QA protocol and the QA protocol previously established for fMRI [13], which includes detecting EPI-related artifacts and monitoring SNR stability. However, in our protocol, the HARDI sequence requires many directions at higher b values, which is more demanding than the fMRI sequence. Furthermore, the SNR calculation methods for these 2 protocols are different: for the fMRI sequence, SNR is calculated from the difference between the sum of the odd and even volumes of time-series images. For HARDI, SNR is calculated from the mean and standard deviation of b0 images. Other multicenter trials have included DW imaging sequences [34, 35] and DTI sequences [15]. Our HARDI sequence differs from DW imaging and DTI sequences in that it uses more gradient directions, which can resolve complex fiber crossings and provide better tractography information.

In conclusion, our initial qualifying scans and phantom QA procedure for a longitudinal HARDI study effectively detected artifacts and were useful for tracking scanner performance. We note that this is not a hypothesis-driven study. However, developing effective quality assurance protocols for multicenter HARDI trials is an ongoing and important problem. We believe that sharing the experiences and quantitative performance metrics from a study such as SPRINT-MS will prove valuable for further advances in multicenter methodology.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank Megan Griffiths, scientific writer for the Imaging Institute, Cleveland Clinic, Cleveland, Ohio, for her editorial assistance.

This research was supported by NIH 1U01NS082329-01A1 (NINDS) and RG 4778-A-6 (National MS Society).

*Magn Reson Imaging*. Author manuscript; available in PMC 2018 January 01.

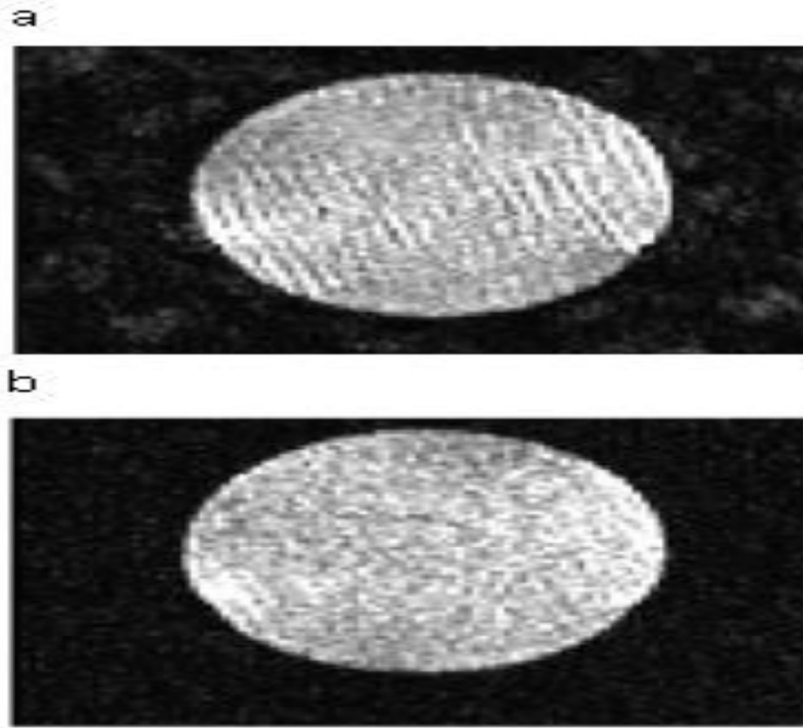
## References

1. Barkhof F, Hulst HE, Drulovic J, Uitdehaag BM, Matsuda K, Landin R, et al. Ibudilast in relapsing-remitting multiple sclerosis: a neuroprotectant? *Neurology*. 2010; 74:1033–40. [PubMed: 20200338]
2. Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magn Reson Med*. 2002; 48:577–82. [PubMed: 12353272]
3. Fox RJ, Beall E, Bhattacharyya P, Chen JT, Sakaie K. Advanced MRI in multiple sclerosis: current status and future challenges. *Neurol Clin*. 2011; 29:357–80. [PubMed: 21439446]
4. Fox RJ, Sakaie K, Lee JC, Debbins JP, Liu Y, Arnold DL, et al. A validation study of multicenter diffusion tensor imaging: reliability of fractional anisotropy and diffusivity values. *AJNR Am J Neuroradiol*. 2012; 33:695–700. [PubMed: 22173748]
5. Teipel SJ, Reuter S, Stieltjes B, Acosta-Cabronero J, Ernemann U, Fellgiebel A, et al. Multicenter stability of diffusion tensor imaging measures: a European clinical and physical phantom study. *Psychiatry Res*. 2011; 194:363–71. [PubMed: 22078796]
6. Takao H, Hayashi N, Kabasawa H, Ohtomo K. Effect of scanner in longitudinal diffusion tensor imaging studies. *Hum Brain Mapp*. 2012; 33:466–77. [PubMed: 21391276]
7. Vollmar C, O’Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, et al. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *Neuroimage*. 2010; 51:1384–94. [PubMed: 20338248]
8. Landman BA, Farrell JA, Huang H, Prince JL, Mori S. Diffusion tensor imaging at low SNR: nonmonotonic behaviors of tensor contrasts. *Magn Reson Imaging*. 2008; 26:790–800. [PubMed: 18499378]
9. Berman JI, Lanza MR, Blaskey L, Edgar JC, Roberts TP. High angular resolution diffusion imaging probabilistic tractography of the auditory radiation. *AJNR Am J Neuroradiol*. 2013; 34:1573–8. [PubMed: 23493892]
10. Chavez S, Storey P, Graham SJ. Robust correction of spike noise: application to diffusion tensor imaging. *Magn Reson Med*. 2009; 62:510–9. [PubMed: 19526513]
11. Zhuang J, Hrabe J, Kangarlu A, Xu D, Bansal R, Branch CA, et al. Correction of eddy-current distortions in diffusion tensor images using the known directions and strengths of diffusion gradients. *J Magn Reson Imaging*. 2006; 24:1188–93. [PubMed: 17024663]
12. Buonocore MH, Gao L. Ghost artifact reduction for echo planar imaging using image phase correction. *Magn Reson Med*. 1997; 38:89–100. [PubMed: 9211384]
13. Friedman L, Glover GH. Report on a multicenter fMRI quality assurance protocol. *J Magn Reson Imaging*. 2006; 23:827–39. [PubMed: 16649196]
14. Zhu T, Hu R, Qiu X, Taylor M, Tso Y, Yiannoutsos C, et al. Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study. *Neuroimage*. 2011; 56:1398–411. [PubMed: 21316471]
15. Walker L, Curry M, Nayak A, Lange N, Pierpaoli C. Brain Development Cooperative Group. A framework for the analysis of phantom data in multicenter diffusion tensor imaging studies. *Hum Brain Mapp*. 2013; 34:2439–54. [PubMed: 22461391]
16. Jones DK, Basser PJ. “Squashing peanuts and smashing pumpkins”: how noise distorts diffusion-weighted MR data. *Magn Reson Med*. 2004; 52:979–93. [PubMed: 15508154]
17. Reese TG, Heid O, Weisskoff RM, Wedeen VJ. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. *Magn Reson Med*. 2003; 49:177–82. [PubMed: 12509835]
18. Stejskal EO, Tanner JE. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time Dependent Field Gradient. *J Chem Phys*. 1965; 42:288–92.
19. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996; 29:162–73. [PubMed: 8812068]
20. Haselgrove JC, Moore JR. Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient. *Magn Reson Med*. 1996; 36:960–4. [PubMed: 8946363]

21. MRI Quality Control Manual. Reston, VA: ACR; 2004. American college of radiology committee on quality assurance in MRI.
22. Pierpaoli C, Walker L, Irfanoglu MO, Barnett AS, Change LC, Koay CG, et al. TORTOISE: An integrated software package for processing of diffusion MRI data. *Proc Intl Soc Mag Reson Med.* 2010; 18:1597.
23. Dietrich O, Heiland S, Sartor K. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. *Magn Reson Med.* 2001; 45:448–53. [PubMed: 11241703]
24. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn Reson Med.* 1995; 34:910–4. [PubMed: 8598820]
25. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage.* 2004; 23(Suppl 1):S208–19. [PubMed: 15501092]
26. Lowe MJ, Koenig KA, Beall EB, Sakaie KE, Stone L, Bermel R, et al. Anatomic connectivity assessed using pathway radial diffusivity is related to functional connectivity in monosynaptic pathways. *Brain Connect.* 2014; 4:558–65. [PubMed: 25117651]
27. Klawiter EC, Schmidt RE, Trinkaus K, Liang HF, Budde MD, Naismith RT, et al. Radial diffusivity predicts demyelination in ex vivo multiple sclerosis spinal cords. *Neuroimage.* 2011; 55:1454–60. [PubMed: 21238597]
28. Lowe MJ, Horenstein C, Hirsch JG, Marrie RA, Stone L, Bhattacharyya PK, et al. Functional pathway-defined MRI diffusion measures reveal increased transverse diffusivity of water in multiple sclerosis. *Neuroimage.* 2006; 32:1127–33. [PubMed: 16798013]
29. Zhou X, Sakaie KE, Debbins JP, Fox RJ, Lowe MJ. Dependence of DTI Measures on SNR in a Multicenter Clinical Trial. *Proc Intl Soc Mag Reson Med.* 2016; 24:3441.
30. Grech-Sollars M, Hales PW, Miyazaki K, Raschke F, Rodriguez D, Wilson M, et al. Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. *NMR Biomed.* 2015; 28:468–85. [PubMed: 25802212]
31. Malyarenko D, Galban CJ, Londy FJ, Meyer CR, Johnson TD, Rehemtulla A, et al. Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. *J Magn Reson Imaging.* 2013; 37:1238–46. [PubMed: 23023785]
32. Tofts PS, Lloyd D, Clark CA, Barker GJ, Parker GJ, McConville P, et al. Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo. *Magn Reson Med.* 2000; 43:368–74. [PubMed: 10725879]
33. Pierpaoli C, Sarlls J, Nevo U, Basser PJ, Horkay F. Polyvinylpyrrolidone (PVP) water solutions as isotropic phantoms for diffusion MRI studies. *Proc Intl Soc Mag Reson Med.* 2009; 17:1414.
34. Chenevert TL, Galban CJ, Ivancevic MK, Rohrer SE, Londy FJ, Kwee TC, et al. Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies. *J Magn Reson Imaging.* 2011; 34:983–7. [PubMed: 21928310]
35. Belli G, Busoni S, Ciccarone A, Coniglio A, Esposito M, Giannelli M, et al. Quality assurance multicenter comparison of different MR scanners for quantitative diffusion-weighted imaging. *J Magn Reson Imaging.* 2016; 43:213–9. [PubMed: 26013043]

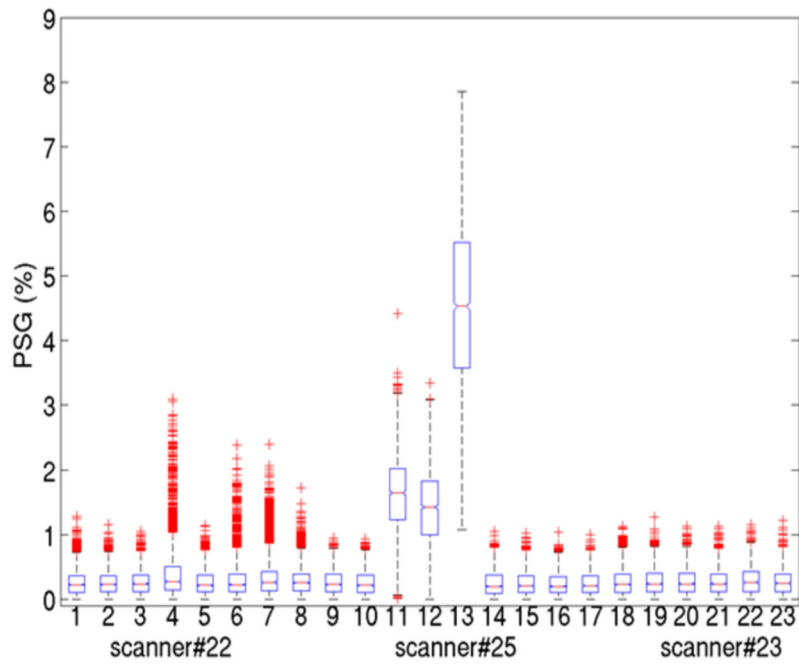


**Fig. 1.** Apparatus for reproducible positioning of the BIRN phantom. Left: adapter ring allowed use of standard holder for cylindrical phantoms for Siemens scanners. Center: detail of adapter rings. Right: GE scanners used standard holders for spherical phantoms.

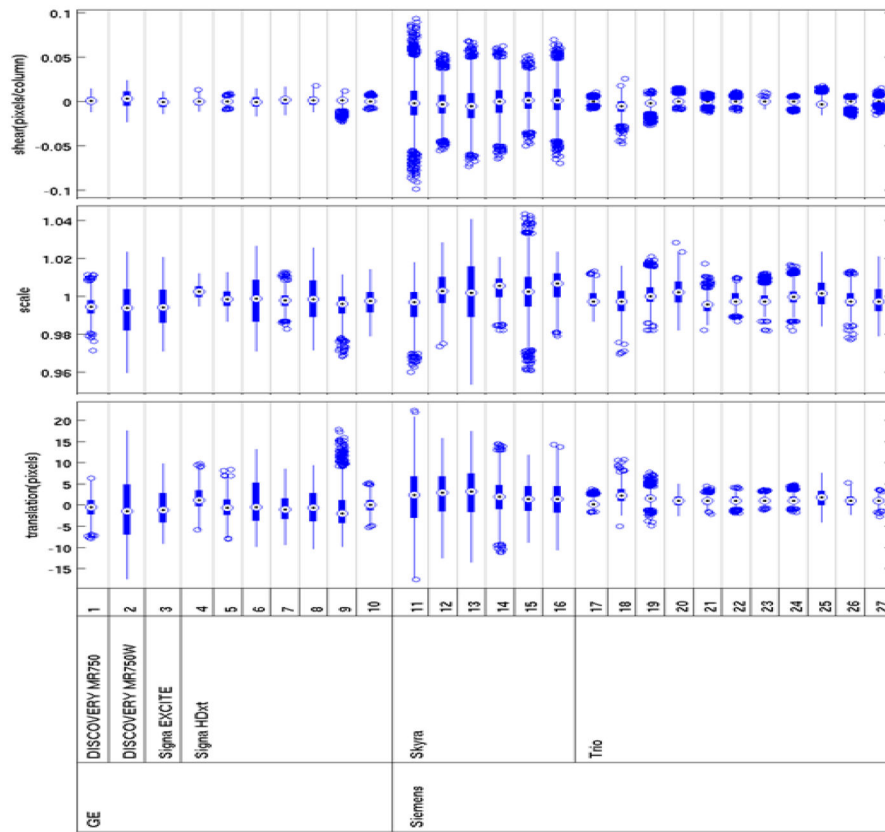


**Fig. 2.** Phantom images with different severities of spiking artifact ranging from severe (a) to minor (b).

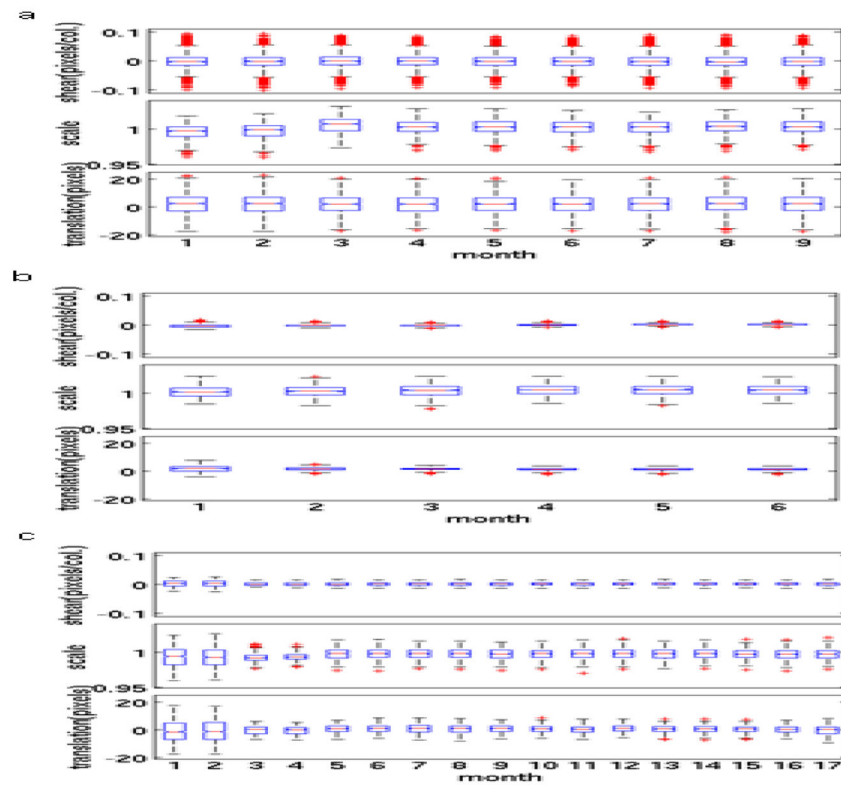




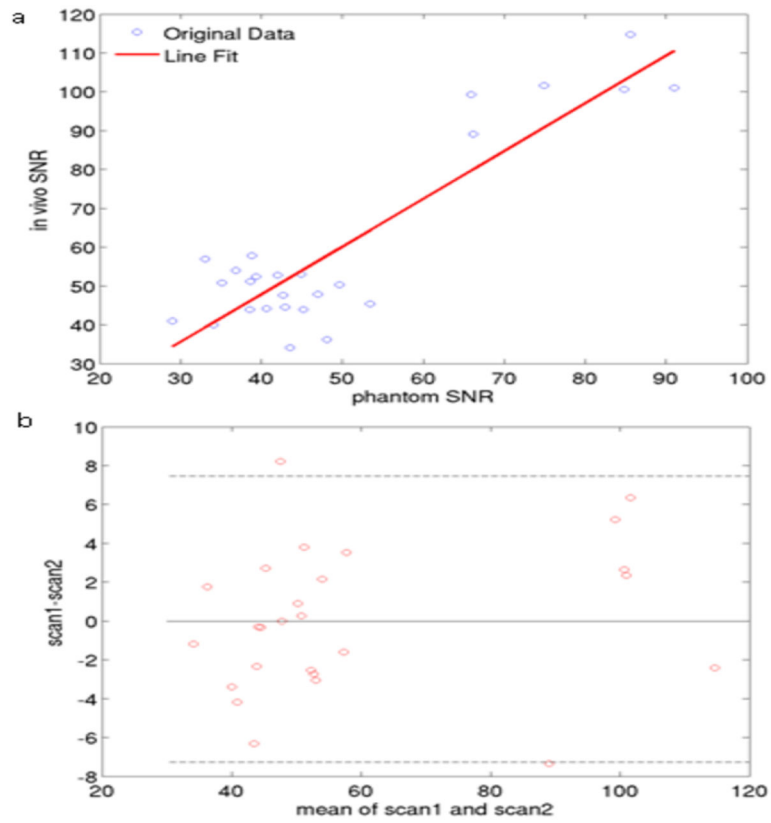
**Fig. 3.** Percent signal ghosting (PSG) over time from 3 scanners. Examples of mild (scanner #22, 1–10), severe (scanner #25, 11–17) and normal (scanner #23, 18–23) levels of ghosting. One box represents one QA dataset for PSG values from all slices and volumes. For each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data values not considered outliers and the outlier points are plotted individually.



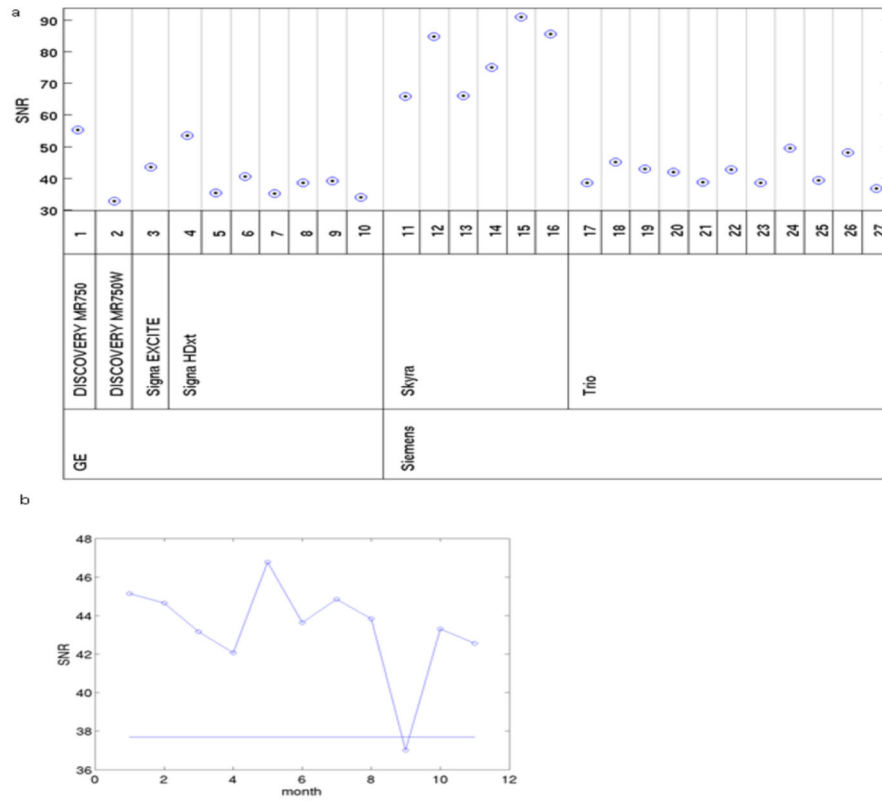
**Fig. 4.** Eddy current distortion parameters indicate values among all slices and diffusion-weighted volumes from qualifying phantom scans. The bulls-eye is the median. The edges of the box are the 25th and 75th percentiles, and the outlier points are plotted individually.



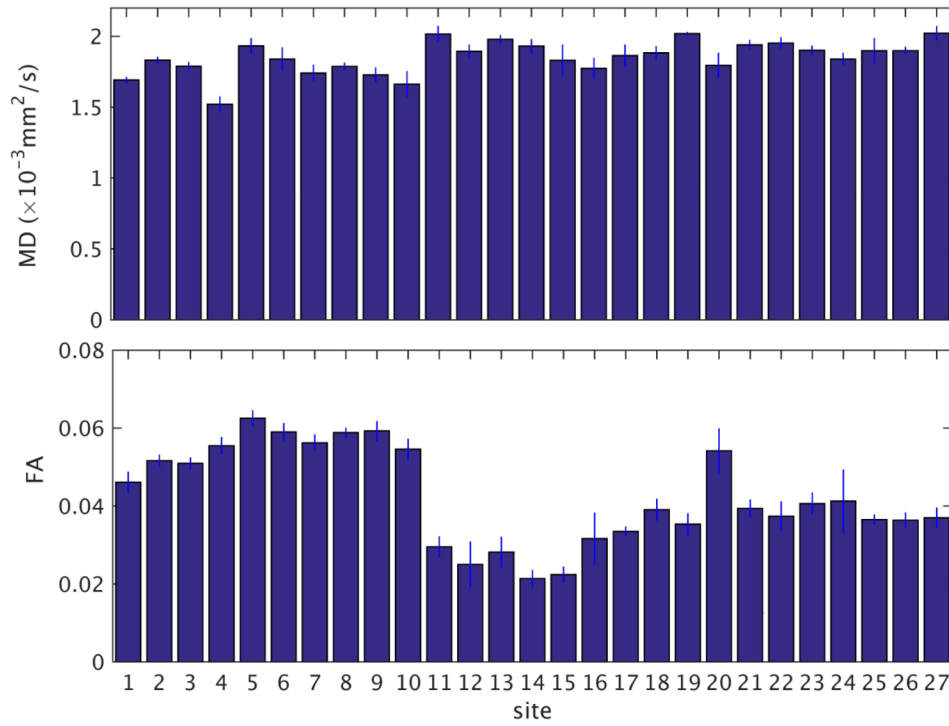
**Fig. 5.** Variation of eddy current distortion parameters over time for 3 selected scanners. (a) Skyra (#11), (b) Trio (#25) and (c) GE Discovery MR750W (#2). In scanner #2, there was a reduction in eddy current distortion between months 2 and 3 associated with a change in head coil. For each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data values not considered outliers, and the outlier points are plotted individually.



**Fig. 6.** (a) Correlation in SNR between phantom and volunteer scans. (b) Volunteer scan-rescan reproducibility using a Bland-Altman plot. The dotted lines represent the 95% confidence intervals.



**Fig. 7.** Comparison of SNR values. (a) SNR values of qualifying scans across 27 MR scanners. (b) Example of variation over time of SNR for one Trio scanner, #17. Straight lines show the thresholds.



**Fig. 8.** Temporal mean and standard deviation of MD and FA for each site. The labeling for the sites is the same as in Fig. 4.

**Table 1**

**3T MR Scanners Involved in SPRINT-MS**

Site	Manufacturer	Model	Software version	Coil
1	GE	DISCOVERY MR750	DV24.0_R01_1344.a	32ch Head/HNS HEAD
2	GE	DISCOVERY MR750W	DV23.1_V02_1317.c	Head 24/8HRBRAIN
3	GE	Signa EXCITE	12.0_M5B_0846.d	8HRBRAIN
4	GE	Signa HDxt	15.0_M4A_0947.a	8HRBRAIN
5	GE	DISCOVERY MR750/Signa HDxt	DV23.1_V02_1317.c/15.0_M4A_0947.a	8HRBRAIN
6	GE	Signa HDxt	HD16.0_V0_1131.a	8HRBRAIN
7	GE	Signa HDxt	HD16.0_V0_1131.a	8HRBRAIN
8	GE	Signa HDxt	HD16.0_V0_1131.a	8HRBRAIN
9	GE	Signa HDxt	HD16.0_V0_1131.a	8HRBRAIN
10	GE	Signa HDxt	HD16.0_V0_1131.a	8HRBRAIN
11	Siemens	Skyra	Syngo MR D13	20-channel head-neck array
12	Siemens	Skyra	Syngo MR D13	20-channel head-neck array
13	Siemens	Skyra	Syngo MR D13	20-channel head-neck array
14	Siemens	Skyra	Syngo MR D13	20-channel head-neck array
15	Siemens	Skyra	Syngo MR D13	20-channel head-neck array
16	Siemens	Skyra	Syngo MR D13	20-channel head-neck array
17	Siemens	Trio1	Syngo MR B17	12-channel standard Siemens
18	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
19	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
20	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
21	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
22	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
23	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
24	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
25	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
26	Siemens	Trio	Syngo MR B17	12-channel standard Siemens
27	Siemens	Trio	Syngo MR B17	12-channel standard Siemens

**Table 2**

## Incidence of Repeated Phantom and Volunteer Scans

Site	Specific issue on volunteer scan	Specific issue on phantom scan
17	Oblique	Wrong phantom position
19	Oblique	
20	Oblique	
3	Motion	
26	Spike	
9	Chemical shift/fat ring artifact	
24	Spike	Spike
6		Wrong slice number
25		Spike

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 3**

## Hardware and Software Changes for Scanners

Site	Scanner model	Scanner change	SNR change
1	GE DISCOVERY MR750	HNS to replace 32-channel head coil	42.4 to 44.6
		Software from 23x to 24x	31.0 to 32.8
2	GE DISCOVERY MR750W	8HRBRAIN to replace Head24 coil with 24x software level	32.8 to 38.2
		Software from 24x to 25x	38.2 to 38.6
7	GE Signa HDxt	Software from 16x to 23x	39.2 to 29.7
9	GE Signa HDxt	New 8HRBRAIN to replace the old coil	50.4 to 50.4
21	Siemens Trio	Gradient coil replacement	39.6 to 39.5
27	Siemens Trio	Hardware and software upgrade to Prisma with VD13	39.9 to 48.6

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript