Left–Right Intensity Asymmetries Vary Depending on Scanner Model for FLAIR and T1 Weighted MRI Images

Arvin Arani, PhD,[1]* Christopher G. Schwarz, PhD,[1] Heather J. Wiste, BA,[2]
Stephen D. Weigand, MS,[2] Petrice M. Cogswell, MD, PhD,[1] Matthew C. Murphy, PhD,[1]
Joshua D. Trzasko, PhD,[1] Jeffrey L. Gunter, PhD,[1] Matthew L. Senjem, MS,[3]
Kiaran P. McGee, PhD,[1] Yunhong Shu, PhD,[1] Matt A. Bernstein, PhD,[1]
John Huston III MD,[1] and Clifford R. Jack Jr MD,[1]
the Alzheimer's Disease Neuroimaging Initiative[†]

Background: Localized regions of left–right image intensity asymmetry (LRIA) were incidentally observed on T_2-weighted (T_2-w) and T_1-weighted (T_1-w) diagnostic magnetic resonance imaging (MRI) images. Suspicion of herpes encephalitis resulted in unnecessary follow-up imaging. A nonbiological imaging artifact that can lead to diagnostic uncertainty was identified.

Purpose: To investigate whether systematic LRIA exist for a range of scanner models and to determine if LRIA can introduce diagnostic uncertainty.

Study Type: A retrospective study using the Alzheimer's Disease Neuroimaging Initiative (ADNI) data base.

Subjects: One thousand seven hundred fifty-three (median age: 72, males/females: 878/875) unique participants with longitudinal data were included.

Field Strength: 3T.

Sequences: T_1-w three-dimensional inversion-recovery spoiled gradient-echo (IR-SPGR) or magnetization-prepared rapid gradient-echo (MP-RAGE) and T_2-w fluid-attenuated inversion recovery (FLAIR) long tau fast spin echo inversion recovery (LT-FSE-IR). Only General Electric, Philips, and Siemens' product sequences were used.

Assessment: LRIA was calculated as the left–right percent difference with respect to the mean intensity from automated anatomical atlas segmented regions. Three neuroradiologists with 37 (**), 32 (**), and 3 (**) years of experience rated the clinical impact of 30 T_2-w three-dimensional FLAIR exams with LRIA to determine the diagnostic uncertainty. Statistical comparisons between retrospective intensity normalized T_1m and original T_1-w images were made.

Statistical Tests: For each image type, a linear mixed effects model was fit using LRIA scores from all scanners, regions, and participants as the outcome and age and sex as predictors. Statistical significance was defined as having a *P*-value <0.05.

Results: LRIA scores were significantly different from zero on most scanners. All clinicians were uncertain or recommended definite diagnostic follow-up in 62.5% of cases with LRIA >10%. Individuals with acute brain pathology or focal neurologic deficits are not enrolled in ADNI; therefore, focal signal abnormalities were considered false positives.

Data Conclusion: LRIA is system specific, systematic, creates diagnostic uncertainty, and impacts IR-SPGR, MP-RAGE, and LT-FSE-IR product sequences.
Level of Evidence: 2
Technical Efficacy Stage: 3

Localized regions of left–right image inhomogeneity occurred on both T_2-weighted (T_2-w) three-dimensional fluid-attenuated inversion recovery (FLAIR) and T_1-weighted (T_1-w) images during a standard diagnostic magnetic resonance imaging (MRI) exam (Fig. 1a) of a Mayo Clinic Study of Aging participant.[1] These images led to suspicion of potential herpes encephalitis that required follow-up (6 days later) imaging (Fig. 1b), which was ultimately determined to be unnecessary, and resulted in avoidable medical concerns for the participant. Throughout this case, the patient did not exhibit the common symptoms of herpes encephalitis, which include headaches, fevers, seizures, loss of consciousness, and confusion,[2] which makes this pathologically unlikely to have been anything but an imaging artifact. What made this even more challenging was that the artifact spanned multiple acquisition types, leading clinicians to interpret it as true underlying pathology, but it was subsequently determined that this might be a systematic imaging issue of which clinicians should be aware of.

In MRI, many factors can contribute to non-tissue-specific image intensity inhomogeneity including static field inhomogeneity (B_o), nonlinearity of gradients, gradient induced eddy currents, and radio-frequency transmission and reception inhomogeneity.[3,4] Efforts to correct MRI intensity inhomogeneity date back to 1986[5,6] and are the subject of previous review articles[7,8]; however, the potential clinical impact of these effects has not been extensively compared across current clinically used MRI systems and models or investigated for neuroimaging research applications. Specifically, intensity comparisons across homologous brain areas are commonly used by neuro-radiologists to identify pathology. If significant nonbiological systematic left–right intensity asymmetries (LRIA) exist, they may lead to diagnostic uncertainty and unnecessary patient burden, as was the case with this initial participant. The experience with the participant described above led to the first aim of this study, which was to determine whether significant systematic LRIA exists for a range of clinical scanner models. Images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu) database were chosen because they include clinical acquisitions from 59 centers worldwide with a large variety of different scanner models. Furthermore, individuals with acute brain pathology or focal neurologic deficits are not enrolled in

FIGURE 1: (a) Index participant's research T_2-w three-dimensional FLAIR three-axis images: a localized region of inhomogeneity with increased T_2 signal (red arrows) was observed in the medial left temporal lobe on three-dimensional FLAIR. In this case, the localized region was suspicious for the potential of herpes encephalitis. (b) Three adjacent axial slices, 6 days later, from the same participant's T_2-w 2D-FLAIR follow-up diagnostic MRI exam that demonstrated no evidence of abnormal signal in the left medial temporal lobe. The participant never exhibited any symptoms of herpes encephalitis strongly suggesting a false positive finding.

the ADNI cohort, and thus focal signal abnormalities would most likely represent artifacts and are extremely likely to be system specific or systematic. Furthermore, the ADNI images that were investigated in this study were obtained from product sequences that would be routinely used by clinicians for diagnostic purposes. The second aim of this study was to measure how often LRIA can introduce diagnostic uncertainty, as determined by a diagnostic review performed by three neuroradiologists.

Materials and Methods

Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu). ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease. This study was approved by institutional review boards and all participants were imaged after obtaining written informed consent.

T_1-w images and T_2-w FLAIR MRIs were obtained from the ADNI phases GO, 2, and 3 (http://www.adni-info.org) databases. All available pairs of 3 tesla (3T) T_2-w FLAIR MRIs and 3T T_1-w MRIs were included. ADNI MRI acquisition protocols have previously been published.[9–11] Atlas-based measurements in 122 regions of interest were initially segmented for each image type. These 122 regions were composed of 48 left and right pairs, where all cerebellar regions were collapsed into a single measurement pair for brevity, and where regions without contralateral pairs were excluded (the pons, dorsal mesopontine, and the vermis). Imaging volumes that did not have valid atlas-based measurements in all 48 regions of interest, those acquired on 1.5T systems, and scanner models with fewer than 10 participants were also excluded. This resulted in a total of 1629 T_2-w three-dimensional FLAIR, 4712 T_2-w 2D FLAIR, and 6107 standard T_1-w image volumes from 974, 1091, and 1753 unique participants, respectively. The volumes were obtained from a total of 17 different scanner models from three manufacturers (General Electric, Siemens, and Philips). Images were downloaded in original de-identified DICOM format and converted to Neuroimaging Informatics Technology Initiative (NIfTI) format.

Product three-dimensional gradient nonlinearity correction was used during online image reconstruction when available and offline in postprocessing where product on-scanner corrections were not available.[4] All images had on-scanner surface coil intensity corrections (General Electric: phased-array uniformity enhancement [PURE], Siemens: Pre-scan normalize, Philips: constant level appearance [CLEAR]) applied prior to any analysis or assessment by radiologists.

To quantify regional LRIA, a standard automated anatomical atlas segmentation approach was used. Atlas regions were localized using advanced normalization tools (ANTs) symmetric normalization[12] with the MCALT T_1 template and ADIR122 atlas. To localize regions in T_2-w FLAIR images, they were rigidly registered to corresponding T_1-w images in SPM12. For each of the 48 regions of interest, a LRIA measure was computed separately for the following

image types: T_2-w three-dimensional FLAIR, T_2-w 2D FLAIR, and T_1-w. The LRIA score was calculated as the left–right percent difference in regional mean intensity (RMI) as:

$$LRIA = \frac{100(RMI_L - RMI_R)}{(RMI_L + RMI_R)/2} \qquad (1)$$

where L/R subscripts correspond to the left and right RMI values, respectively.

To evaluate if LRIA could be reduced in postprocessing, an addition inhomogeneity correction was applied to the product direct-from-scanner T_1-w MRI images; these are referred to as T_1m images. Intensity inhomogeneity and tissue-class probabilities were estimated by unified segmentation[13] in SPM12 with population-optimized tissue priors and settings from the Mayo Clinic Adult Lifespan Template (MCALT; https://www.nitrc.org/projects/mcalt/).[14] This pipeline uses the N3 algorithm for intensity homogeneity,[15] followed by additional correction using unified segmentation[13] in SPM5. The same regional parcellations from the above SPM12/ANTs-based pipeline were used for estimating regional intensity values from these SPM5-based N3m images.

To identify if LRIA impacted diagnostic uncertainty, three neuroradiologists were presented with product on-scanner T_2-w three-dimensional FLAIR image volumes and asked them to fill out a short questionnaire for each image based on their interpretation of intensities in the hippocampus. The calculated hippocampal LRIA scores were used to select 30 participants' T_2-w three-dimensional FLAIR volumes evenly distributed between the following five absolute percent difference bins: 1) <3%, 2) >4% and <6%, 3) >10% and <12%, 4) >16% and <18%, and 5) >22% such that hippocampal LRIA scores ranged between −22% and 33%. An even distribution of all three vendor types was also selected. Three board-certified neuroradiologists with 37 (**), 32 (**), and 3 (**) years post-fellowship experience were independently shown image volumes, given specific instructions to primarily focus on asymmetry in the hippocampus, and asked to score each volume from 1 to 5, where 1 represented no asymmetry and 5 represented severe asymmetry, and additionally asked to state which side had greater intensity. For each case, the radiologist was also asked "Is clinical follow-up required?" and given the choice between "Definitely not," "Uncertain," and "Definitely yes" as possible responses. The two senior radiologists were aware that the data came from the ADNI database, while the third radiologist was completely blinded by the origin of the scans.

Statistical Analysis

Differences in the distributions of age, sex, and clinical diagnosis across individuals with T_2-w three-dimensional FLAIR, T_2-w 2D FLAIR, and T_1-w images were tested using generalized estimating equations to account for repeated measures across image types within an individual. For each sequence, boxplots were made of the LRIA values across all participants, scans, and scanner models.

Linear mixed effects models were used to help investigate the extent of LRIA across scanners and regions. For each imaging type, a linear mixed effects model was fit using LRIA scores from all scanners, regions, and participants as the outcome and age and sex as predictors, treated as fixed effects in the model. Scanner, region, and

participant were treated as random effects. The estimated difference in mean (95% confidence interval [CI]) LRIA for a 10-year difference in age and for male vs. female sex were summarized for each imaging type. Additionally, the estimated mean (95% CI) LRIA for the scanner random effect summarizes the overall LRIA bias on each scanner, while using shrinkage estimation to account for multiple comparisons.

Similar mixed effects models to the ones described above were also fit separately for three different scanners (one from each vendor) and separately for the T_1-w and T_2-w three-dimensional FLAIR sequences. Because these models were fit separately by scanner, only random intercepts for region and participant were included. Estimated mean (95% CI) LRIA in each region was summarized.

A variance-components analysis was performed within each image type using the linear mixed effects models described above and the magnitudes of the standard deviations (SDs) of the three variance components (scanner, region, and participant) plus the SD for unexplained variance were compared.

All statistical analyses were done using the R language and environment for statistical computing version 4.0.3 (R Foundation for Statistical Computing). Mixed effects models were fit using the lme4 package. Two-sided P-values less than 0.05 were considered statistically significant.

Results

Among the 1753 participants with a T_1-w scan, 41% were cognitively unimpaired, 44% had MCI, and 15% had dementia. The median age was 72 (range 55 to 95) and 878 (50%) were male (Table 1). There were small differences in median age between the subsets with T_2-w three-dimensional FLAIR or T_2-w 2D FLAIR compared to the overall cohort with T_1-w scans (74 and 73 vs. 72). The subset with T_2-w three-dimensional FLAIR included more females, while the subset with T_2-w 2D FLAIR included fewer females compared to the overall sample with T_1-w scans (53% and 47% vs. 50%). Participants with T_2-w three-dimensional FLAIR were more likely to be cognitively unimpaired compared to the overall sample with T_1-w scans (55% vs. 41%), while the participants with T_2-w 2D FLAIR were more likely to have MCI or dementia (67% vs. 59%). A detailed summary of the number of scans for each scanner model is outlined in Table 2.

TABLE 1. Characteristics of Participants with T_2-w Three-Dimensional FLAIR, T_2-w Two-Dimensional FLAIR, T_1, and T_1m SEQUENCES. Data Are summarized for the First Scan Date Within a Person. All Individuals Had T_1 and T_1m Sequences. A Subset Had T_2-w Three-Dimensional FLAIR and/or T_2-w Two-Dimensional FLAIR Sequences

	3D FLAIR (N = 974)	2D FLAIR (N = 1091)	T_1/T_1m (N = 1753)
Age, years			
Median (Q1, Q3)	74 (68, 80)	73 (68, 78)	72 (67, 78)
Range	55–97	55–95	55–95
Sex			
Female	514 (53%)	516 (47%)	875 (50%)
Male	460 (47%)	575 (53%)	878 (50%)
Clinical diagnosis			
Cognitively			
unimpaired	537 (55%)	358 (33%)	719 (41%)
Mild cognitive			
impairment	322 (33%)	547 (50%)	768 (44%)
Dementia	109 (11%)	181 (17%)	256 (15%)

Q1 = first quartile or 25th percentile; Q3 = third quartile or 75th percentile.

TABLE 2. Scanner MODEL with the number of Scans and Number of Unique Participants in Parentheses Used in the Analysis for T_2-w Three-Dimensional FLAIR, T_2-w Two-DIMENSIONAL FLAIR, T_1-w, and T_1m sequences

Model #	3D FLAIR	2D FLAIR	T_1-w/T_1m
1	260 (140)	682 (171)	907 (255)
2	93 (54)	31 (27)	124 (75)
3	0	29 (10)	27 (10)
4	32 (27)	0	30 (27)
5	0	468 (131)	462 (131)
6	65 (43)	545 (137)	570 (162)
7	98 (64)	18 (14)	109 (68)
8	0	35 (15)	35 (15)
9	96 (60)	100 (28)	186 (82)
10	0	20 (12)	18 (12)
11	8 (6)	210 (44)	204 (46)
12	7 (7)	16 (11)	21 (16)
13	192 (119)	15 (13)	203 (121)
14	472 (279)	20 (20)	490 (291)
15	117 (78)	356 (102)	452 (160)
16	82 (52)	1501 (349)	1512 (391)
17	107 (68)	666 (165)	757 (222)
Total no. of scans (participants):	1629 (974)	4712 (1091)	6107 (1753)

The LRIA score for each of the image types across all participants, scans, and regions, within each scanner model have been plotted in Fig. 2. Box plots not centered at zero indicate evidence of systematic asymmetries and the mixed effects model analysis confirmed that most scanners had estimated mean LRIA scores that were significantly different than 0 (Table 3). Varying degrees of systematic intensity biases in LRIA scores are present across the 17 scanners models and across the four different image types. More specifically, in the T_2-w three-dimensional FLAIR analysis, seven scanners had significant L > R bias with mean LRIA ranging from 1.7% to 6.2%, two scanners had significant R > L bias with mean LRIA ranging from −1.6% to −5.1%, and four scanners had

no significant bias ($P \geq 0.06$). In the T_2-w 2D FLAIR analysis, six scanners had significant L > R bias ranging from 0.9% to 4.6%, six scanners had significant R > L bias ranging from −1.5% to −3.0%, and four scanners had no significant bias ($P \geq 0.05$). For the T_1-w analysis, 4 scanners had significant L > R bias ranging from 1.3% to 3.4%, 10 scanners had significant R > L bias ranging from −1.0% to −5.4%, and 3 scanners had no significant bias ($P \geq 0.10$).

T_1m processing generally contained the LRIA scores to within 10% as can be seen in Fig. 2. For T_1m, only 0.08%, or 239 out of 293,136 observations (6107 scans × 48 regions), had LRIA values greater than 10%. In contrast, for T_1-w 10%, or 29,751 observations, had LRIA values greater
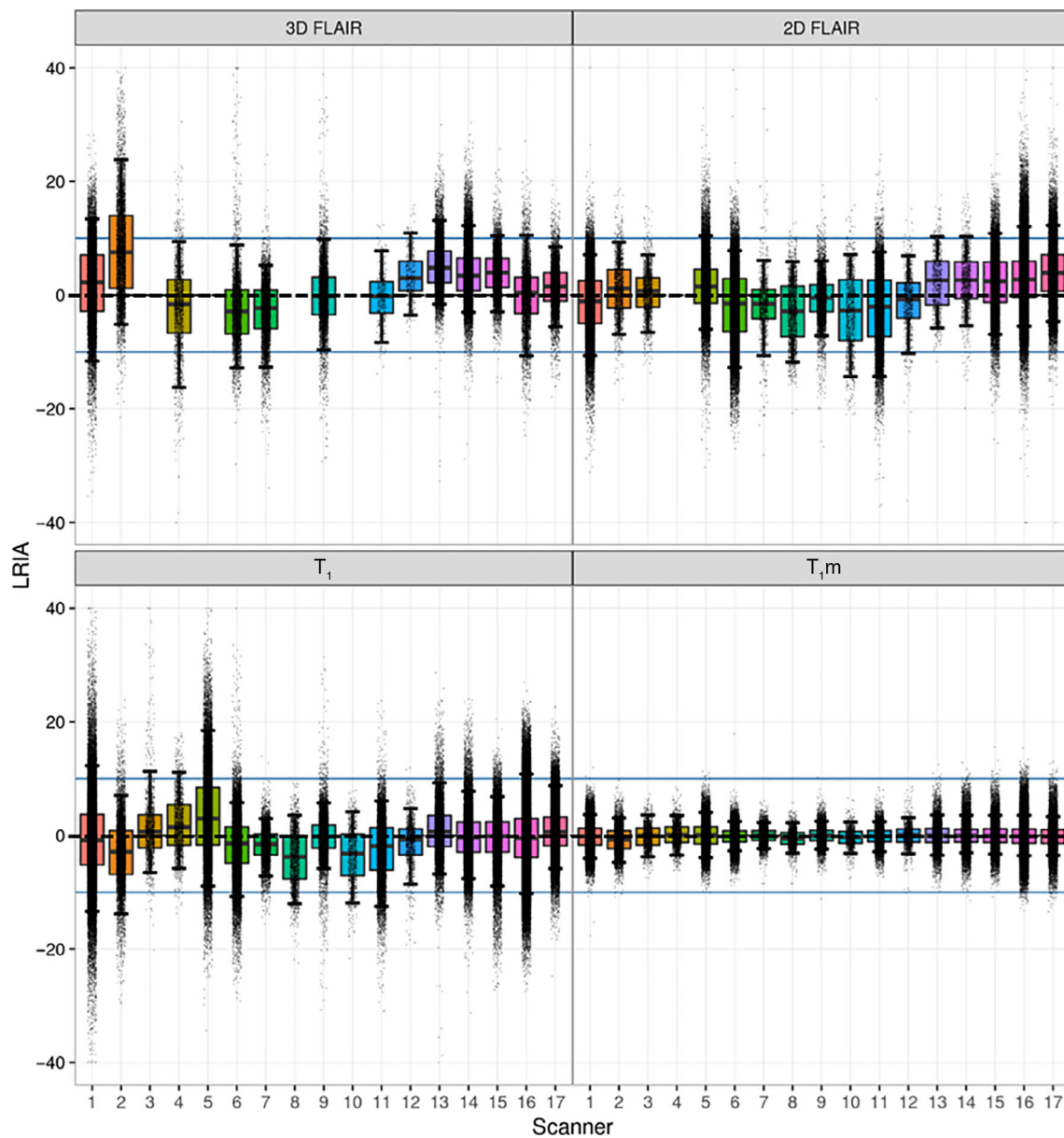

FIGURE 2: Plot of LRIA scores across all participants, scans, and regions within each of the 17 scanner models. Each of the 48 regional LRIA scores per scan has been represented as a black dot and a few points have been truncated at ±40%. The boxes represent the 1st quartile (25th percentile), median (50th percentile) and 3rd quartiles (75th percentile), and the whiskers represent the 5th and 95th percentiles.

TABLE 3. Estimated Mean (95% CI) Difference in LRIA for Fixed Effects of Sex and Age and for Random Effects of Scanner Model. The Age and Sex Estimates Are Small (Less Than 0.5% of the Observed Effects)

	3D FLAIR		2D FLAIR		T$_1$-w		T$_1$m	
	Estimate (95% CI)	P	Estimate (95% CI)	P	Estimate (95% CI)	P	Estimate (95% CI)	P
Fixed effects								
Male vs. female	0.4% (0.0% to 0.8%)	0.04	0.1% (−0.1% to 0.3%)	0.49	−0.1% (−0.4% to 0.1%)	0.32	0.04% (0.01% to 0.07%)	0.006
Age, per 10 years	−0.4% (−0.7% to −0.2%)	<0.001	0.2% (0.1% to 0.3%)	0.004	0.5% (0.4% to 0.6%)	<0.001	0.01% (−0.01% to 0.02%)	0.50
Random scanner effects								
Scanner 1	2.0% (1.3% to 2.6%)	<0.001	−1.5% (−2.1% to −1.0%)	<0.001	1.3% (0.5% to 2.1%)	0.001	−0.2% (−0.3% to −0.1%)	0.002
Scanner 2	6.2% (5.3% to 7.0%)	<0.001	0.6% (−0.0% to 1.2%)	0.05	−5.4% (−6.2% to −4.5%)	<0.001	−0.8% (−1.0% to −0.7%)	<0.001
Scanner 3			0.5% (−0.2% to 1.2%)	0.13	1.7% (0.8% to 2.5%)	<0.001	−0.3% (−0.4% to −0.1%)	<0.001
Scanner 4	−1.6% (−2.4% to −0.7%)	<0.001			−1.2% (−2.1% to −0.3%)	0.005	0.1% (−0.1% to 0.2%)	0.23
Scanner 5			1.4% (0.9% to 2.0%)	<0.001	3.4% (2.6% to 4.2%)	<0.001	−0.0% (−0.1% to 0.1%)	0.61
Scanner 6	−5.1% (−5.9% to −4.3%)	<0.001	−2.2% (−2.8% to −1.6%)	<0.001	−2.2% (−3.0% to −1.3%)	<0.001	−0.2% (−0.3% to −0.1%)	0.001
Scanner 7	−0.7% (−1.5% to 0.0%)	0.06	−3.0% (−3.7% to −2.3%)	<0.001	−1.9% (−2.7% to −1.0%)	<0.001	−0.1% (−0.2% to 0.1%)	0.36
Scanner 8			−2.7% (−3.7% to −1.7%)	<0.001	−3.5% (−4.7% to −2.2%)	<0.001	−0.5% (−0.7% to −0.3%)	<0.001
Scanner 9	0.1% (−0.8% to 1.0%)	0.84	0.4% (−0.2% to 1.1%)	0.19	0.4% (−0.5% to 1.2%)	0.40	−0.1% (−0.2% to 0.1%)	0.31
Scanner 10			−2.1% (−3.1% to −1.1%)	<0.001	−2.5% (−3.8% to −1.2%)	<0.001	−0.3% (−0.5% to −0.1%)	0.001
Scanner 11	0.0% (−2.3% to 2.3%)	0.99	−2.7% (−3.5% to −2.0%)	<0.001	−5.0% (−5.9% to −4.2%)	<0.001	−0.2% (−0.3% to −0.1%)	0.005
Scanner 12	3.2% (1.0% to 5.3%)	0.003	−0.9% (−1.9% to 0.1%)	0.06	−1.5% (−2.8% to −0.3%)	0.02	−0.0% (−0.2% to 0.2%)	0.84
Scanner 13	5.0% (4.3% to 5.7%)	<0.001	0.9% (0.2% to 1.6%)	0.008	0.7% (−0.2% to 1.5%)	0.10	0.1% (−0.1% to 0.2%)	0.29
Scanner 14	3.7% (3.1% to 4.3%)	<0.001	2.4% (1.7% to 3.0%)	<0.001	−1.0% (−1.8% to −0.2%)	0.01	0.0% (−0.1% to 0.1%)	0.59
Scanner 15	3.8% (3.0% to 4.6%)	<0.001	2.1% (1.6% to 2.7%)	<0.001	−1.8% (−2.6% to −1.0%)	<0.001	−0.1% (−0.2% to 0.1%)	0.30
Scanner 16	0.3% (−0.6% to 1.1%)	0.49	2.5% (2.0% to 3.1%)	<0.001	−0.7% (−1.5% to 0.1%)	0.10	−0.1% (−0.2% to 0.0%)	0.05
Scanner 17	1.7% (0.8% to 2.5%)	<0.001	4.6% (4.0% to 5.1%)	<0.001	1.6% (0.8% to 2.4%)	<0.001	−0.2% (−0.3% to −0.1%)	<0.001

than 10%. In the T_1m analysis, nine scanners had no significant bias ($P \geq 0.05$), while eight scanners had significant, but unlikely to be clinically important, R > L bias with mean LRIA values ranging from −0.2% to −0.8% (Table 3). The directionality of the bias was often the same for T_2-w 2D and three-dimensional FLAIR images (9 of 12 scanners) and T_2-w 2D FLAIR and T_1-w images (11 of 16 scanners), but closer to parity between the T_2-w three-dimensional FLAIR and T1-w images (7 of 13 scanners).

An example of the variation in the spatial pattern and directionality of the LRIA scores across anatomical regions in the brain for three scanner models from three vendors has been plotted in Fig. 3. Interestingly, T_2-w three-dimensional FLAIR biases occurred in opposite directions for a majority of the anatomical locations for scanner 7 as compared to scanners 2 and 14. In contrast, T_1 images generally had biases going in the same direction for these three different vendors. However, to some extent, the same regions generally have high absolute LRIA scores across vendors.

Figure 4 summarizes the variance components analysis findings. The normalized density functions graphically summarize variances due to scanner model, anatomical region, participant, and unexplained error, for all four image types.

For T_2-w FLAIR (three-dimensional and two-dimensional) and T_1-w images, a significant proportion of the variance is due to scanner model (21% of total variance for three-dimensional FLAIR, 14% for 2D FLAIR, and 13% for T_1-w). However, the T_1m results show that when homogeneity correction is applied (in this case retrospectively), 1% of the variance in the data was explained by scanner model. In contrast, in the T_1m model 42% of the total variability was explained by expected biological intensity variations in different anatomical regions compared to 31% in T_1-w, 6% in three-dimensional FLAIR, and 11% in 2D FLAIR. In other words, the T_1m data analysis becomes sensitive to variations in anatomy, which is desirable and expected, as opposed to systematic biases due to scanner type, and the specific participant being examined. Less than 0.5% of the observed effects could be contributed to age and sex (Table 3).

A sensitivity analysis was done where the models were fit with clinical diagnosis added as a fixed effect. Random effect estimates by scanner were essentially unchanged and the variance components were the same as shown in Fig. 4. While there were some significant differences in mean LRIA for MCI compared to cognitively unimpaired (T_2-w two-dimensional FLAIR) or for participants with dementia
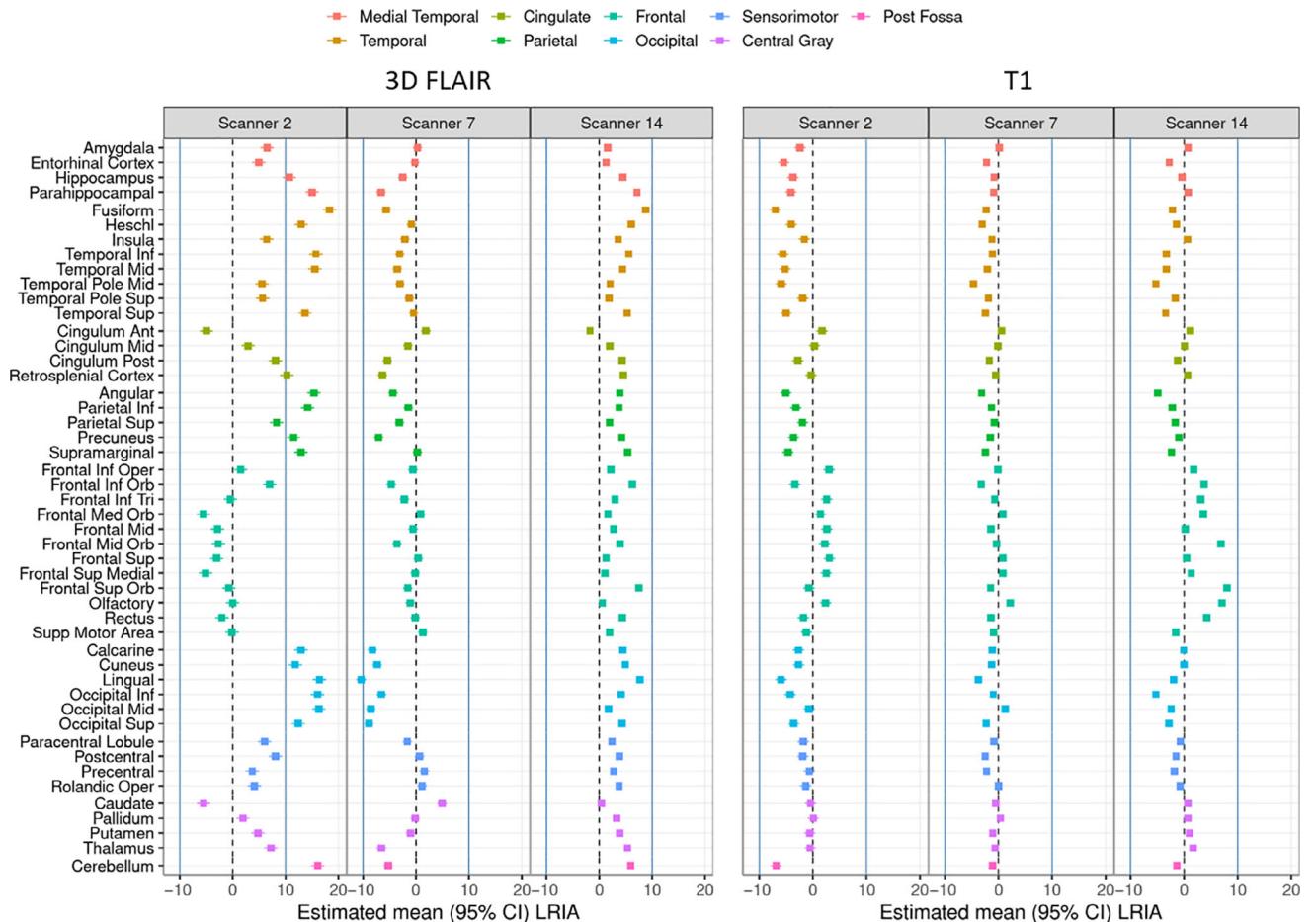

FIGURE 3: Plot of the mean regional LRIA with 95% CIs for three exemplar scanners. Estimates are from the random region-specific intercepts in the linear mixed effects model fit within scanner and imaging type.
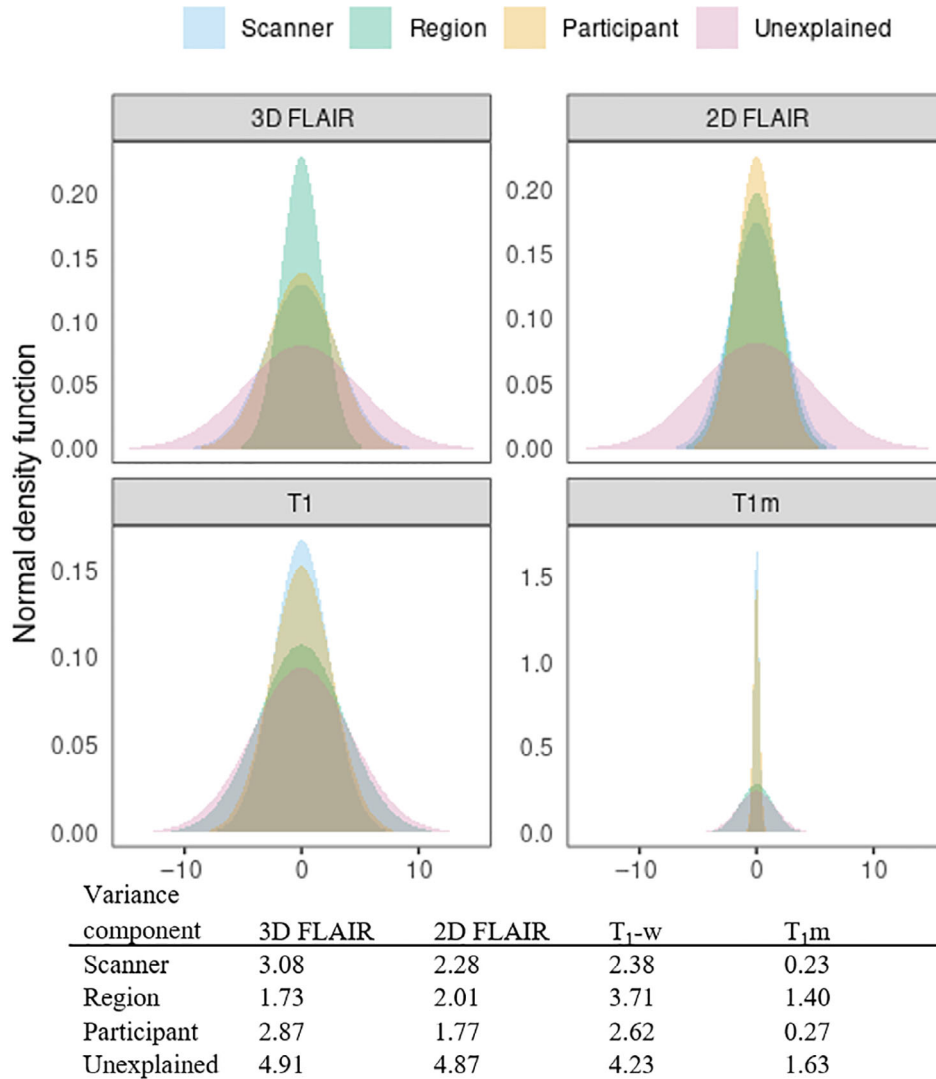
Variance component	3D FLAIR	2D FLAIR	T_1-w	T_1m
Scanner	3.08	2.28	2.38	0.23
Region	1.73	2.01	3.71	1.40
Participant	2.87	1.77	2.62	0.27
Unexplained	4.91	4.87	4.23	1.63

FIGURE 4: Normal densities and estimated LRIA SDs from a variance components analysis accounting for participant, region, scanner, and unexplained error within each modality.

compared to cognitively unimpaired (T_2-w three-dimensional FLAIR, T_1-w, T_1m) across the four image types, the effect sizes were small (0.5% or less).

The radiologists' assessments of the diagnostic impact of LRIA on product on-scanner images and the reported follow-up recommendations are summarized in Fig. 5. All clinicians were uncertain or would recommend definite diagnostic follow-up in over 43% of cases (reviewer 1: 13/30, reviewer 2: 14/30, reviewer 3: 21/30, with consensus on 7 cases), with 62.5% (reviewer 1: 8/13, reviewer 2: 11/14, reviewer 3: 11/21) of these cases occurring with absolute LRIA scores above 10%. To illustrate the potential impact and likelihood of a 10% LRIA score, for two diagnostically important brain regions, the median LRIA across participants within scanner model are shown in Fig. 6 by region and image type. Scanner models 2, 8, and 13 had absolute median LRIA above 10% for at least one of the two regions and image types. Many scanners had median LRIA scores between 5% and 10%,

particularly for the fusiform region. This suggests a high prevalence of LRIA in the fusiform and hippocampus regions and also demonstrates that the polarity of the LRIA value can change depending on the scanner model.

Discussion

LRIAs can mimic disease and create diagnostic uncertainty that can lead to unnecessary additional testing and patient burden. Statistically significant, and clinically important, LRIA scores were found across participants within specific scanner models for both T_1-w and T_2-w FLAIR images (three-dimensional and two-dimensional) across all three manufacturers. The polarity of the left–right asymmetry varied across scanner models and for some vendors also varied across sequences on the same system.

The systematic cross-participant behavior of LRIA suggests that this is unlikely to be associated with a specific
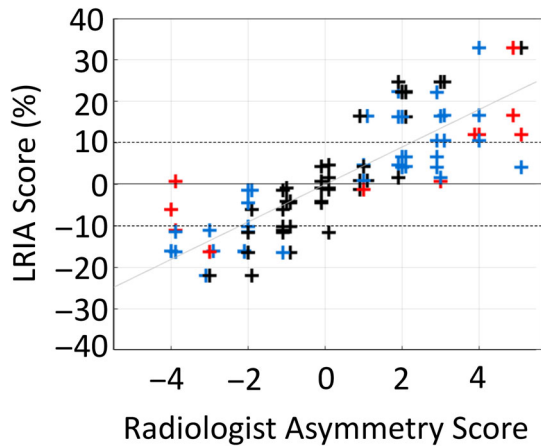
FIGURE 5: Three experienced radiologist's asymmetry scores (assessment localized to the hippocampus) are plotted against the LRIA score (percent difference in intensity measured in the hippocampus, where left side > right side) with automated atlas-based segmentation on T_2-w three-dimensional FLAIR volumes. The same 30 image volumes from 30 different ADNI participants were scored. The color of the data points corresponds to the response of each radiologist to the question: "Is clinical follow-up required?" where: + (black) Definitely not, + (blue) Uncertain, + (red) Definitely yes.

susceptibility artifact, but rather a combination of systematic features such as B1 receive-field correction limitations[7] or radio-frequency penetration artifacts.[16] Small effect sizes in relation to age, sex, and diagnosis were observed in comparison to scanner effect sizes. This further supports the

conclusion that there is a systematic nonbiological property to the LRIA artifact.

Vendors have addressed intensity inhomogeneity by implementing B1 receive-field corrections that rely on low-resolution maps. Commercially, these corrections are referred to as CLEAR, PURE, and prescan normalization. Furthermore, vendors also provide an on-scanner nonlinear gradient correction option. This correction includes an intensity correction step (Jacobian brightness correction), which causes areas of the image that spread out (increase in volume) to decrease in intensity.[17] Both of these product on-scanner corrections were already applied to all images prior to image assessment by the radiologists in this study, and significant LRIA was still observed.

The retrospective nonparametric, nonuniform intensity normalization that was used to produce the T_1m images demonstrated the capability of improving LRIA across all models. In Fig. 4, it can be seen that the variation due to scanner and participant was reduced in T_1m-processed images, while expected regional intensity variations were maintained. Post-processing signal intensity corrections are already commonly used prior to any statistical analysis[18] of the images for research studies; these methods are reviewed in Vovk et al.[3] Although these methods are effective at reducing biases in morphological metrics,[18] they are not currently available as commercial products to assist clinicians during routine clinical exams. Furthermore, these postprocessing approaches also
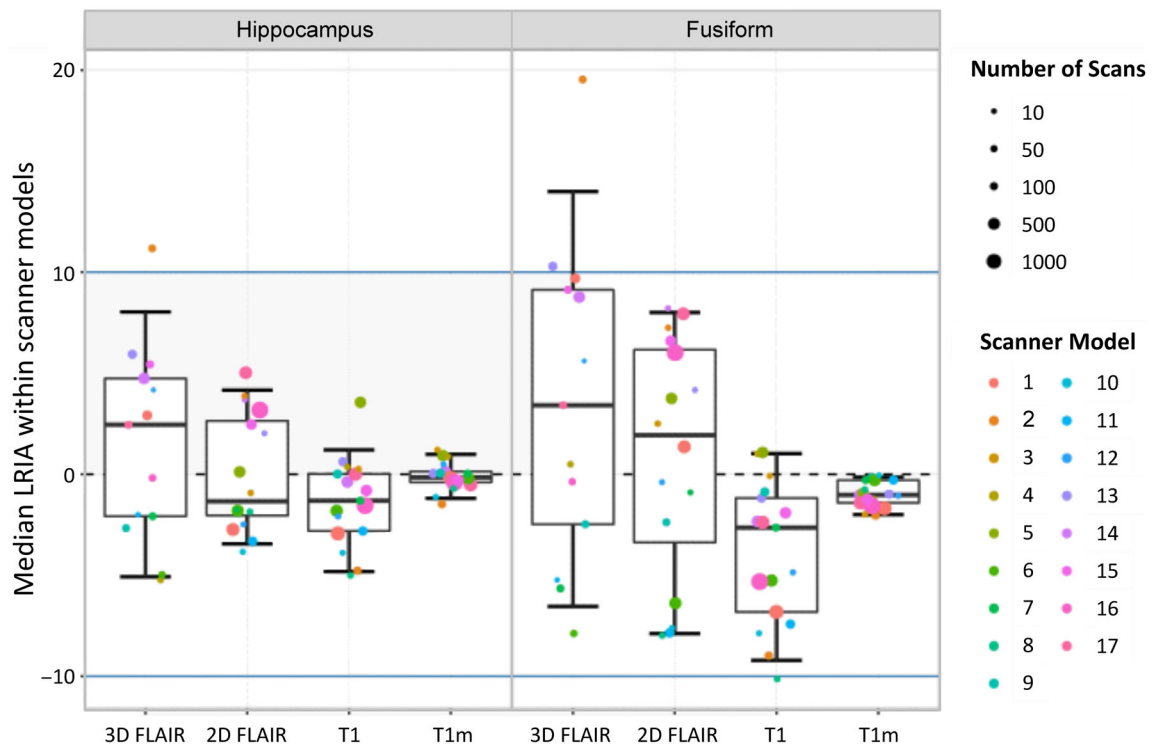

FIGURE 6: Plot of the median LRIA across participants within each scanner model and image type for the hippocampus and fusiform regions of interest. Each point represents the median LRIA for a scanner in that region. The color of the point indicates scanner model and the size of the point indicates number of scans contributing to the estimated median.

have a fundamental flaw in that they could potentially remove or attenuate some true pathology because they are not modeled to correct the physical phenomenon producing these artifacts, but rather any general intensity inhomogeneity. Furthermore, since eight scanners still maintained a significant intensity bias, post correction, this suggests that global low-order inhomogeneity corrections may not be sufficient to eliminate higher-order localized left–right asymmetries, further emphasizing the importance of accurate physics-based models to correct for these changes. To accurately correct these artifacts, true raw data (k-space from each coil element and all associated calibration files) would most likely be needed, which is rarely saved in neuroimaging research, let alone in clinical settings. Vendor involvement and on-scanner strategies would be the most practical solution to this problem.

In addition to receive coil sensitivity, spatial variations in the B1 transmit fields and the geometry[16] of the head may also contribute to uneven contrast. Although parallel excitation using dedicated coil setups and simultaneous RF transmission on multiple channels was demonstrated to be feasible in 2005,[19] it has not been translated into routine clinical use and body transmit coils are the current standard. Until such technologies are adopted, it may be helpful to use pads with high dielectric constants to increase the radio-frequency wavelengths and limit the localized signal fluctuations in routine clinical practice.[20]

New hardware and software that enables on-scanner inhomogeneity corrections could help address many of these issues. Further development and implementation directly by vendors are recommended. In parallel, lower tolerances in scanner calibrations could potentially be an immediate remedy for reducing the severity of LRIA. Currently, the American College of Radiology[21] recommends that during quality control assessment, a percent integral uniformity (PIU) of greater than 82.0% should be maintained, and calibration failure is reported at 80%. A PIU of 80% translates to a percent different of roughly 40%, as percent difference (i.e., LRIA score) is defined in this manuscript. From the radiologists' visual and diagnostic assessments, it can be expected that LRIA scores of 10% could introduce clinical uncertainty, and thus the current ACR standard of 40% is well above what radiologists can observe by eye. Furthermore, rather than comparing left–right pairs, the PIU compares the highest and lowest intensity regions in the entire image from a uniform calibration phantom and generally does not look at three-dimensional volume imaging. For certain scanner models and regions, the average LRIA scores approach $\pm 10\%$; therefore, this recommendation may need to be adjusted. Another alternative approach is for the field to start incorporating more quantitative imaging maps into clinical practice, where absolute T_1 and T_2 values are obtained rather than just weighted images.

Limitations

The original images where left–right asymmetries created diagnostic uncertainty were prospectively observed; however, a key limitation of this manuscript is that the rest of the analysis was all done on a retrospectively acquired dataset. Secondly, some scanner models had a limited number of participants and could not be included in the study. Lastly, on-scanner corrections capable of removing the observed left–right asymmetries were not demonstrated. These corrections fell outside the scope of this study but need to address in future investigations. These limitations could have introduced unintentional inherent bias into the results.

Conclusion

In conclusion, until on-scanner improvements are implemented and/or new quality control guidelines are in place, it is important that the radiology community be aware that systematic left–right intensity asymmetries are capable of mimicking disease and should be considered when interpreting images to help avoid unnecessary additional testing and patient burden. For institutions with a small range of scanners, this phenomenon may be somewhat predictive in the sense that clinicians can self-calibrate to systematic asymmetries in the data. In institutions with multiple vendors and MRI models, the severity and directionality of the expected bias could be characterized for each individual scanner and this information could be provided to clinicians during their review.

References

1. Roberts RO, Geda YE, Knopman DS, et al. The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology 2008;30(1):58-69.

2. Sili U, Kaya A, Mert A. Herpes simplex virus encephalitis: Clinical manifestations, diagnosis and outcome in 106 adult patients. J Clin Virol 2014;60(2):112-118.

3. Vovk U, Pernus F, Likar B. A review of methods for correction of intensity inhomogeneity in MRI. IEEE Trans Med Imaging 2007;26(3):405-421.

4. Jovicich J, Czanner S, Greve D, et al. Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. Neuroimage 2006;30(2):436-443.

5. Haselgrove J, Prammer M. An algorithm for compensation of surface-coil images for sensitivity of the surface coil. Magn Reson Imaging 1986;4(6):469-472.

6. McVeigh ER, Bronskill MJ, Henkelman RM. Phase and sensitivity of receiver coils in magnetic resonance imaging. Med Phys 1986;13(6):806-814.

7. Bernstein MA, Huston J III, Ward HA. Imaging artifacts at 3.0T. J Magn Reson Imaging 2006;24(4):735-746.

8. Jones RW, Witte RJ. Signal intensity artifacts in clinical MR imaging. Radiographics 2000;20(3):893-901.

9. Jack CR, Bernstein MA, Borowski BJ, et al. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. Alzheimers Dement 2010;6(3):212-220.

10. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27(4):685-691.

11. Team TAMC. ADNI MRI SCANNER PROTOCOLS. Volume 2022. WWW Document. Available from: http://adni.loni.usc.edu/methods/documents/mri-protocols/

12. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 2008;12(1):26-41.

13. Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005;26(3):839-851.

14. Schwarz CG, Gunter JL, Ward CP, et al. [P2–415]: The Mayo Clinic Adult lifespan template: Better quantification across the lifespan. Alzheimers Dement 2017;13(7S_Part_16):P792.

15. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17(1):87-97.

16. Sled JG, Pike GB. Standing-wave and RF penetration artifacts caused by elliptic geometry: An electrodynamic analysis of MRI. IEEE Trans Med Imaging 1998;17(4):653-662.

17. Tao S, Trzasko JD, Shu Y, Huston J III, Bernstein MA. Integrated image reconstruction and gradient nonlinearity correction. Magn Reson Med 2015;74(4):1019-1031.

18. Yushkevich PA, Avants BB, Das SR, Pluta J, Altinay M, Craige C. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3 T MRI data. Neuroimage 2010;50(2):434-445.

19. Ullmann P, Junge S, Wick M, Seifert F, Ruhm W, Hennig J. Experimental analysis of parallel excitation using dedicated coil setups and simultaneous RF transmission on multiple channels. Magn Reson Med 2005;54(4):994-1001.

20. Foo TKF, Hayes CE, Kang Y-W. Reduction of RF penetration effects in high field imaging. Magn Reson Med 1992;23(2):287-301.

21. Radiology ACo. American College of Radiology MRI Accreditation Program. Volume 2021; 2018. WWW Document. Available from: https://www.acraccreditation.org/-/media/ACRAccreditation/Documents/MRI/LargePhantomGuidance.pdf