Analyzing Deep Learning Techniques in Accelerated Clinical Brain Magnetic Resonance Imaging for Multiple Sclerosis

by

Ashika Mani

BS Statistics, University of Pittsburgh, 2020

Submitted to the Graduate Faculty of the

Department of Biostatistics in the

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Ashika Mani

It was defended on

April 26, 2021

and approved by

Committee Member: Jeanine Buchanich, PhD, MPH, MEd, Research Associate Professor, Biostatistics, Graduate School of Public Health, University of Pittsburgh

Committee Member: Jenna Carlson, PhD, Assistant Professor, Biostatistics, Graduate School of Public Health, University of Pittsburgh

Committee Member: Zongqi Xia, MD, PhD, Assistant Professor, Neurology and Biomedical Informatics, School of Medicine, University of Pittsburgh

Thesis Advisor: Ada Youk, PhD, Associate Professor, Biostatistics, Graduate School of Public Health, University of Pittsburgh

Analyzing Deep Learning Techniques in Accelerated Clinical Brain Magnetic Resonance Imaging for Multiple Sclerosis

Ashika Mani, MS

University of Pittsburgh, 2021

Abstract

Background: Magnetic resonance imaging (MRI) scans are routine clinical procedures for monitoring people with multiple sclerosis (MS). Accelerated MRI scan time is motivated by patient discomfort, timely scheduling, and financial burden associated with conventional MRI scans.

Objective: We examined the application of a deep learning (DL) model in restoring the image quality of accelerated clinical brain MRI scans for MS.

Methods: We acquired fast 3D T1w BRAVO and fast 3D T2 FLAIR MRI sequences alongside conventional scans. Using a subset of the scans, we trained the DL model to generate images from fast scans with quality similar to the conventional scans and then applied the model to the remaining scans. We calculated clinically relevant T1w volumetrics (normalized brain volume, normalized thalamic volume, normalized gray matter volume, and normalized white matter volume) for all scans. We performed paired t-tests for conventional, fast, and fast with DL for these volumetrics, and fit repeated measures linear mixed-effects models to test for differences in correlations between volumetrics and clinically relevant patient-reported outcomes. We performed equivalence tests to compare fast scans with DL and conventional scans to examine equivalence in image quality as well as equivalence in association with patient-reported clinical outcomes.

Results: We found statistically significant differences between conventional scans and fast scans with DL for all T1w volumetrics. There was no difference in the extent to which the key volumetrics and clinical outcomes are correlated between fast scans with DL and conventional scans, but there was not sufficient evidence to prove that the correlations were equivalent.

Conclusion: There is currently no evidence to support that fast scans with DL produce images of equivalent quality to conventional scans. However, fast scans with DL have the potential to inform clinically relevant outcomes in MS.

Public health significance: Limited research has been done regarding the application of deep learning models to improve the image quality of accelerated scans in clinical brain MRIs for people with multiple sclerosis. The results of this analysis can inform practitioners as to how to further incorporate and improve on MRIs utilizing deep learning.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

Multiple Sclerosis (MS) is a chronic neurological disease that affects 2.3 million people worldwide and causes symptoms such as loss of vision, loss of balance, fatigue, memory and concentration problems, among others[1]. Routine magnetic resonance imaging (MRI) scans are vital to monitoring disease activity and progress in those with MS[2]. Generally, people with MS undergo brain MRI scans every six months to 2 years. Most people undergo MRI scans yearly, with factors such as changes to treatment and disease severity changing the scan frequency. Longer MRI scan times contribute to patient discomfort, increased motion and reduced image quality, delays in scheduling and potentially high medical cost— which is a significant driver of the financial burden for patients with MS[3]. Therefore, accelerating the MRI scans could benefit those with MS and generally improve access to essential diagnostic imaging. However, these accelerated MRI scans pose a challenge due to diminished image quality (*e.g.,* contrast to noise ratio, resolution). To our knowledge, there has been no research on comparing these accelerated MRI scans to conventional MRI scans to see if they produce equivalent images.

Artificial intelligence approaches seek to address the loss of MRI quality in accelerated scans. Deep learning (DL) models, such as convolutional neural networks (CNNs), enhance MRI quality of the accelerated scans without compromising relevant image information passed through each layer[4-7]. Deep learning has been utilized in a variety of imaging applications, such as brain segmentation, stroke imaging, breast cancer, imaging in oncology, and medical ultrasounds[8]. Research on CNNs is rapidly increasing and is considered to be pioneered by Yang et. al. at the 2016 Conference on Neural Information Processing Systems. In recent years, CNNs have been utilized in MRI scans for MS, but are still not widely used in clinical practice[9]. CNNs show

comparable capability in regards to MS lesion segmentation, which is paramount to accurately diagnosing and developing a treatment plan in those with MS, and often greater adaptability to the inherent class imbalance that comes with some of the most successful supervised learning algorithms such as Random Forests (a supervised segmentation method based on voxel-wise classifiers)[10]. Class imbalance refers to issues that arise in the training datasets, particularly in medical image classification where certain types of diseases only appear in a small portion of the dataset. Deep back-projection network (DBPN) is a class of CNN that utilizes an iterative up- and down- sampling layers, providing a mechanism to aid in self-correction of errors using back-projection[11]. We are not aware of an application of DBPN or any other deep leaning models on MRI scans in the MS population thus far. In this study, we evaluated the clinical application of a DL model based on DBPN that employed noise-reducing and sharpness-enhancing functions.

Normalized brain volume, normalized thalamic volume, normalized gray matter volume, and normalized white matter volume are all known to inform neurological disability in MS. Clinically relevant patient reported outcomes such as Patient Determined Disease Steps (PDDS) and Multiple Sclerosis Rating Scale-Revised (MSRS-R) are common measures to assess gait impairment and symptomatic burden, respectively. There have not been previous studies that have explored the correlation between these clinical outcomes with brain volume measurements. It is known that brain volume correlates with and predicts future disability[12]. This premise motivates the importance of understanding the relationship between the correlations of volumetric measures and clinical outcomes between conventional scans and accelerated MRI scans that utilize deep learning methods.

This thesis aims to assess MRI scans enhanced with DL to see if they produce images of equivalent quality in the MS population by comparing normalized volumetric measures.

Additionally, we will assess whether the DL model improves the quality of accelerated MRI scans to the extent that the key volumetrics preserve their correlation with clinically relevant neurological outcomes comparable to the benchmark conventional MRI scans.

2.0 Methods

2.1 Data Source

Participants were recruited from a clinic-based, prospective MS cohort study (Prospective Investigation of Multiple Sclerosis in the Three Rivers Region, PROMOTE) based in the Pittsburgh region. The Institutional Review Board of the University of Pittsburgh approved this study and all participants completed the informed consent process.

2.2 MRI Acquisition

One hundred fifteen participants underwent routine clinical brain MRI studies on a GE Discovery MR750 3T scanner between September 2018 and January 2020. In addition to the institutional protocol that includes the standard (or conventional) 3D T1-weighted (T1w) BRAVO (FE/PE/SE: 220x220x126, scan time 2:57), 3D T2 FLAIR (FE/PE/SE: 256x224x240, scan time 6:40) and other routine clinical sequences, we acquired an accelerated (or fast) 3D T1w BRAVO (FE/PE/SE: 220x128x64, scan time 1:13) and an accelerated (or fast) 3D T2 FLAIR (FE/PE/SE: 256x128x120, scan time 2:17) within the same MRI exam.

We obtained the results from the deep learning (DL) model based on DBPN[11] to enhance the image quality for the fast sequences through the use of the software SubtleMR™ (https://subtlemedical.com/usa/subtlemr/). The DL model input the fast sequences and generated high resolution images similar to that of the conventional sequences. The output of the DL model

had twice the slices as that of the input. We trained the DL model with the first 15 scans, with images from the conventional sequences. A L1 loss was applied in training to measure the difference between the DL output and conventional scans. The L1 loss function is also known as the least absolute deviations (LAD), and it minimizes the sum of the absolute differences between the target value and the estimated values. We applied image pre-processing, including image registration[13] , bias field correction[14], and image normalization to the training data. We implemented the proposed DL model in TensorFlow and trained on an NVIDIA V100 GPU with an ADAM optimizer[15]. After training, we applied the DL model to the remaining 100 scans for evaluation.

2.3 MRI Analysis

The overall workflow is shown in Figure 1. T1w volumetric measures were prioritized for this study. First, we examined raw T1w images for: (1) voluntary or involuntary patient movement during image acquisition that could decrease the accuracy of the scan (4 out of 108); (2) acquisitions in the wrong phase encoding orientation (magnetic gradient field applied right to left rather than left to right) (7 out of 108); and (3) missing scans (1 out of 108). Nine patients participated in a second MRI during the duration of the study. We used the Freesurfer software version 6.0 (http://surfer.nmr.mgh.harvard.edu/) and then estimated the volumes of 96 sets of T1w MR images. Volumetric analysis included the following regions: total brain volume, total thalamus, total cerebral gray matter, total cerebral white matter and intracranial volumes. These regions make up the MRI metrics that will be discussed later. We then extracted the volumes from

the automatic segmentation file “aseg” and then normalized them by the individual intracranial volume. All of the normalized volumes are unitless.

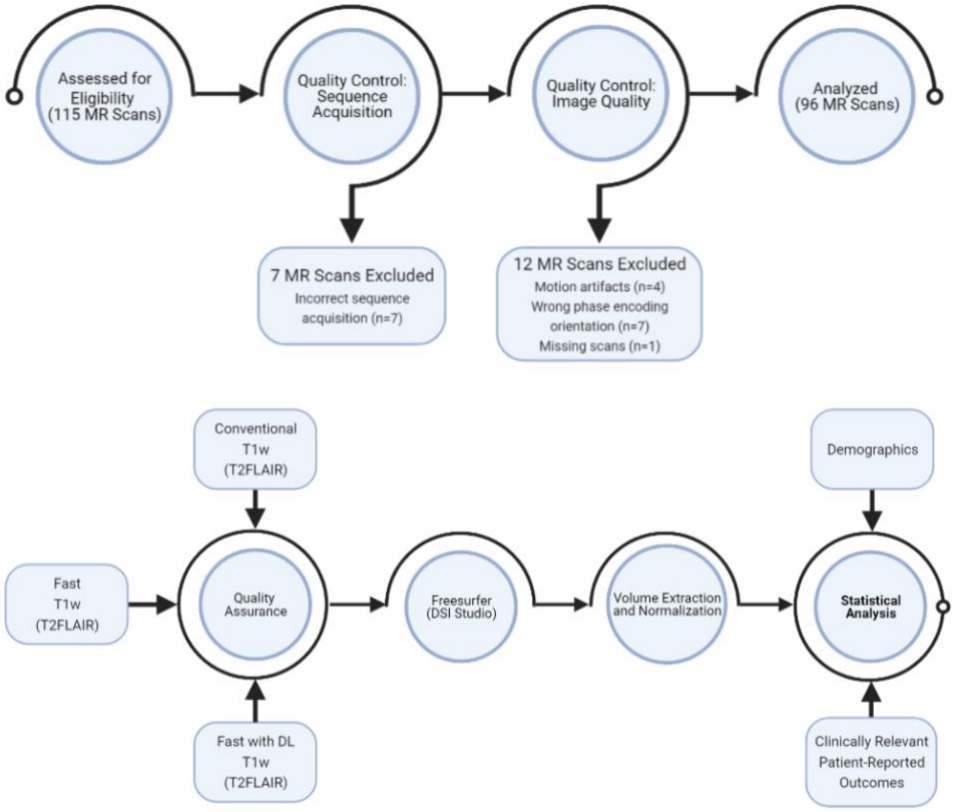


Figure 1 Flowchart summarizing the image processing method

2.4 Covariates

The variables utilized for analysis included age at first scan, sex, race, ethnicity, disease duration in years, relapsing form of multiple sclerosis (RMS) status, standard-efficacy treatment status, and high-efficacy treatment status. Sex refers to the genotypical sex of the patient (male or

female). Race information was collected for all patients and people were classified as Caucasian if they were of European origin. Patients were classified into two categories for ethnicity, Non-Hispanic and Hispanic, where Hispanic includes anyone from a Spanish-speaking background. Disease duration refers to the length of time between the date of the first symptoms to the date of the first MRI. RMS includes people with relapsing-remitting MS (RRMS) and clinically-isolated syndrome (CIS). Standard-efficacy treatment included the drugs Aubagio, Avonex, Betaseron, Copaxone/glatopa, Extavia, Gilenya, Mayzent, Novandrone, Plegridy, Rebif, Tecfidera, Vumerity, Zeposia, and Zinbryta. High-efficacy treatment included the drugs Lemtrada, Mavenclas, Ocrevus, Rituxan, and Tysabri. We also used two clinically relevant patient-reported outcomes of neurological function in people with Multiple Sclerosis. The first is Patient Determined Disease Steps (PDDS) which assessed the gait impairment, ranging from 0 to 8 defined by the North American Research Consortium on Multiple Sclerosis (NARCOMS) Registry. They define the levels of the PDDS scale in order as: normal, mild disability, moderate disability, gait disability, early cane, late cane, bilateral support, wheelchair/scooter, and bedridden[16-18]. The second is the Multiple Sclerosis Rating Scale-Revised (MSRS-R), which assessed the MS symptom burden across eight domains: walking, using arms and hands, vision, speech, swallowing, cognition, sensation, and bowel and bladder control[19,20]. Each domain had a sub-score ranging from 0 (no symptoms) to 4 (severe disability) for a maximum total score of 32. Interviews were conducted through the online platform PatientsLikeMe[21].

2.5 Statistical Analysis

All analyses were completed using R version 4.0.3[22]. For all tests, a two-sided P-value less than 0.0125 was indicative of statistical significance as we used the Bonferroni Correction method to obtain this significance level (0.05/4 = 0.0125), given the four different volumetric measures being tested. For descriptive variables, we expressed continuous data as mean and standard deviation (SD) or medians and interquartile ranges, and categorical data as frequencies and percentages.

2.5.1 Paired T-Tests

In order to assess the impact of scan type on volume measure, we first performed paired t-tests to pairwise compare the three types of MRI scans (conventional, fast, fast with DL) for the four volumetric measures: normalized brain volume (NBV), normalized thalamic volume (NThV), normalized gray matter volume (NGMV) and normalized white matter volume (NWMV). To help quantify the changes and compare them between the four volumetrics we calculated the mean percentage difference between each MRI method. This was calculated for each of the four volumetric measures by taking the mean volumetric measure for the first method and subtracting the mean volumetric measure of second method and dividing it by the second method *100.

2.5.2 Paired TOST Equivalence Tests

In order to examine equivalence in image quality, we performed equivalence tests to compare conventional MRI scans and fast scans with DL for the same four MRI metrics:

normalized brain volume, normalized thalamic volume, normalized gray matter volume, and normalized white matter volume. Equivalence tests are recommended when the goal is to compare different conditions, which are type of scans in this case. Traditional difference-based tests such as t-tests are used to make claims about differences in population means. Generally, if the null hypothesis is not rejected in these tests, researchers use that to conclude equivalence of population means. When using these traditional tests, equivalence of population means will usually be found when studies are under-powered and thus it is suggested to use tests of equivalence[23]. We used the two one-sided tests (TOST) approach for equivalence testing, first proposed by Schuirmann[24]. This method involves decomposing the null hypothesis and alternative hypothesis into two sets of one-sided hypotheses. The TOST procedure can be used to statistically reject the presence of effects large enough to be considered worthwhile[25]. Because the two variables of interest were measured on the same subject, we utilized a paired version of this test. The null hypothesis (H_0) and alternative hypothesis (H_1) then become:

$$H_0: \mu_D \leq -\varepsilon \; or \; \mu_D \geq \varepsilon$$

$$H_1: -\varepsilon < \mu_D < \varepsilon$$

Equation 1

μ_D represents the mean of the differences between the two variables and ε represents the margin of equivalence. The package 'equivalence' in R was used to perform the TOST equivalence tests[26]. Normally, the margin of equivalence is chosen based on a known prior clinically relevant value. We did not have that in this case, so we utilized the putative placebo strategy proposed by Wiens[27]. This putative placebo strategy involves using a decimal fraction of the mean difference between the active control (conventional scan) and putative placebo (fast scan) for the equivalence margin. As our aim was to compare the test condition (fast with DL) with the active control, we

utilized a .1 decimal fraction of the mean difference between the active control and the putative placebo to maintain a conservative estimate of the true difference. Fast scans with DL would be equivalent to conventional scans if they differed by less than 10% of the observed difference between fast scans and conventional scans.

2.5.3 Repeated Measures Linear Mixed-Effects Models

To measure the association between clinical measures PDDS and MSRS-R and the four MRI metrics, we developed separate repeated measures multivariate linear mixed-effects models using R and the package 'lme4'[28]. A random subject effect was included in the models with the type of MRI conducted as the between-subject factor. The four respective volume measurements for the MRI metrics formed the dependent variables. We included an interaction term between the clinical measures (PDDS and MSRS-R) tested and the type of MRI as a fixed effect in our model, with the conventional MRI type as the reference point. The interaction terms quantified (1) differences in associations with clinical measures between fast scans and conventional scans, and (2) differences in associations with clinical outcomes between fast scans with DL and conventional scans. We also included age, sex, race/ethnicity, disease duration in years, clinical type, standard-efficacy treatment status, and high-efficacy treatment status as fixed effects in the model, to measure the contribution on the association between the two clinical metrics and four MRI metrics. For each of the MRI metrics, we created models with two clinical measures as outcomes: PDDS and MSRS-R.

$$volume_{ij} = \beta_0 + \beta_1 type_j + \beta_2 PDDS_i + \beta_3 type_j * PDDS_i + age_i + sex_i + nonhiswhite_i + diseasedur_i + RMS_i + stan_i + high_i + \zeta_i + \epsilon_{ij}$$

Equation 2

$$volume_{ij} = \beta_0 + \beta_1 type_j + \beta_2 MSRS_i + \beta_3 type_j * MSRS_i + age_i + sex_i + nonhiswhite_i + diseasedur_i + RMS_i + stan_i + high_i + \zeta_i + \epsilon_{ij}$$

Equation 3

Where:

$volume_{ij}$ = brain volume measurement for subject i with scan type j

j = 1 Conventional Scan

j = 2 Fast scan

j = 3 Fast scan with DL

$PDDS_i$ = PDDS for subject i

$MSRS_i$ = MSRS-R for subject i

age_i = age (at first scan) in years for subject i

sex_i = sex (1 if female, 0 if male) for subject i

$nonhiswhite_i$ = race/ethnicity (1 if Non-Hispanic European descent, 0 if not)

$diseasedur_i$ = disease duration (in years) for subject i

RMS_i = clinical type of MS (1 if relapsing clinical type, 0 if not) for subject i

$stan_i$ = standard-efficacy treatment status (1 if standard treatment, 0 if not)

$high_i$ = high-efficacy treatment status (1 if standard treatment, 0 if not)

ζ_i = random intercept for subject i

ϵ_{ij} = error term for subject i with scan type j

Assuming:

$\zeta_i \sim \mathrm{N}(0, \psi)$

$\epsilon_{ij} \sim N(0, \theta)$

Predicted residuals and random intercepts were plotted to check normality according to the model assumptions.

2.5.4 Multivariate TOST Equivalence Tests on Interaction Terms

Once the linear mixed-effects models were obtained, we compared the associations with patient-reported clinical outcomes between conventional MRIs and fast scans with DL. We were interested in applying an equivalence test on the interaction terms between the clinical outcomes and the type of scan to see if the associations themselves were equivalent. We used the TOST approach, developing a test that simultaneously tested both null hypotheses that state that the estimated difference between the mean of fast scans with DL and conventional is $\leq -\varepsilon$ *or* $\geq \varepsilon$ based on equivalence tests for linear regression models proposed by Mascha[29].

$$H_0: \mu_3 - \mu_1 \leq -\varepsilon \; or \; \mu_3 - \mu_1 \geq \varepsilon$$
$$H_1: -\varepsilon < \mu_3 - \mu_1 < \varepsilon$$

Equation 4

In this case, $\mu_3 - \mu_1$ represents the mean difference between fast scans with DL volume and conventional scans volume relative to the clinical outcome. Unlike equivalence tests, noninferiority tests aim to test that the fast scan with DL perform no worse than the conventional scan. The hypotheses null hypothesis (H_0) and alternative hypothesis (H_1) for a noninferiority test then become

$$H_0: \mu_3 - \mu_1 \leq -\varepsilon$$
$$H_1: \mu_3 - \mu_1 > -\varepsilon$$

Equation 5

The test statistic for a non-inferiority test follows below:

$$\frac{\widehat{\mu_3} - \widehat{\mu_1} + \varepsilon}{\sqrt{S_p(\frac{1}{n_3} + \frac{1}{n_1})}} \text{ where } S_p = \sqrt{\frac{(n_3 - 1)s_3^2 + (n_1 - 1)s_1^2}{n_3 + n_1 - 2}} \qquad \textbf{Equation 6}$$

It is important to note that the equivalence test is the intersection of two non-inferiority trials, thus resulting in the following test statistics:

$$\frac{\widehat{\mu_3} - \widehat{\mu_1} + \varepsilon}{\sqrt{S_p(\frac{1}{n_3} + \frac{1}{n_1})}} \text{ and } \frac{\widehat{\mu_3} - \widehat{\mu_1} - \varepsilon}{\sqrt{S_p(\frac{1}{n_3} + \frac{1}{n_1})}} \qquad \textbf{Equation 7}$$

In order to claim equivalence we must test the first test statistic to see if it is $> T_{n_3+n_1-2,1-\alpha}$ and the second test statistic to see if it is $< T_{n_3+n_1-2,1-\alpha}$ Further reducing the test statistics, the denominators simply become the estimated standard error of the difference and the numerators represent the coefficients in the regression model to become:

$$\frac{\widehat{\beta_3} + \varepsilon}{\widehat{SE}_{\widehat{\beta_3}}} > T_{n_3+n_1-2,1-\alpha} \text{ and } \frac{\widehat{\beta_3} - \varepsilon}{\widehat{SE}_{\widehat{\beta_3}}} < T_{n_3+n_1-2,1-\alpha} \qquad \textbf{Equation 8}$$

β_3 represents the coefficient for the interaction term between the clinical outcome and scan type from the linear mixed-effects model that we generated. If either test fails to reject the null hypothesis, we fail to reject the equivalence test. If both reject the null hypothesis, we can conclude that they are equivalent. Again, because we had no known prior clinical correlation, we used a fraction of .1 of the fixed effect estimate between fast scans and conventional scans for the equivalence margin. We tested for equivalence between fast scans with DL and conventional scans within the linear mixed-effects model framework.

3.0 Results

3.1 Study Sample

After the 19 scans with quality control failure were excluded, this study had 87 unique patients with Multiple Sclerosis. Table 1 displays the demographic characteristics of these patients. There were a total of 96 MRI scans. Nine patients had two MRIs occurring on separate days. The mean age of this population was 47 years. Most participants were women (70.2%) and of Non-Hispanic European descent (80.5%). Most participants (86.2%) had the relapsing clinical type of MS. The mean Patient Disease Determined Steps (PDDS) was 1.7 and the median Multiple Sclerosis Rating Scale-Revised (MSRS-R) was 4, with mostly mild physical disability and symptoms. The mean disease duration (interval between the date of MRI and date of first symptoms) in years was 15 years. Most participants (72.4%) received some treatment at the time of the MRI, while 5.7% of participants received high-efficacy treatment. In those who received treatment, only one treatment regimen was utilized through the course of this study.

Table 1 Patient Characteristics

	Total (N=87)
Age (years), Mean ± SD	46.8 ± 13.3
Men, n (%)	26 (29.8)
European-descent, n (%)	73 (83.9)
Non-Hispanic, n (%)	84 (96.6)
Non-Hispanic European descent, n (%)	70 (80.5)
PDDS, Mean ± SD	1.7 ± 1.8
MSRS-R, Median (IQR)	4 (2-9)
RMS, n (%)	75 (86.2)
Disease Duration (years), Mean ± SD	14.9 ± 19.3

No Treatment, n (%)	24 (27.6)
High-efficacy Treatment, n (%)	5 (5.7)

3.1.1 MRI Volumetric Measures

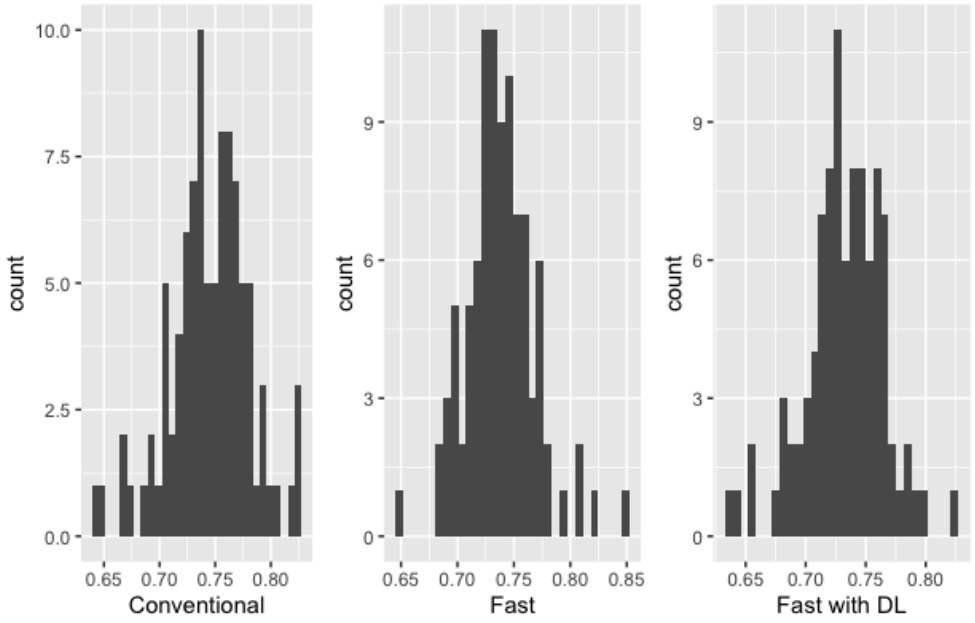
Table 2 shows the mean and variability of the various MRI metrics. Due to normalization, all volumes have no unit. Normalized brain volume averaged approximately 0.74 using conventional scans (sd = 0.036) and fast scans (sd = 0.031). Average normalized brain volume was slightly lower using fast scans with DL at about 0.73. Normalized thalamic volume averaged about 0.009 (sd = 0.001) for all three MRI methods. Normalized gray matter volume had the highest average in conventional scans, followed by fast scans with deep learning (DL), and fast scans. Normalized white matter volume had the highest average in fast scans, followed by fast scans with DL and conventional scans.

Table 2 Summary Statistics (mean ± SD) of MRI Metrics by each MRI Method

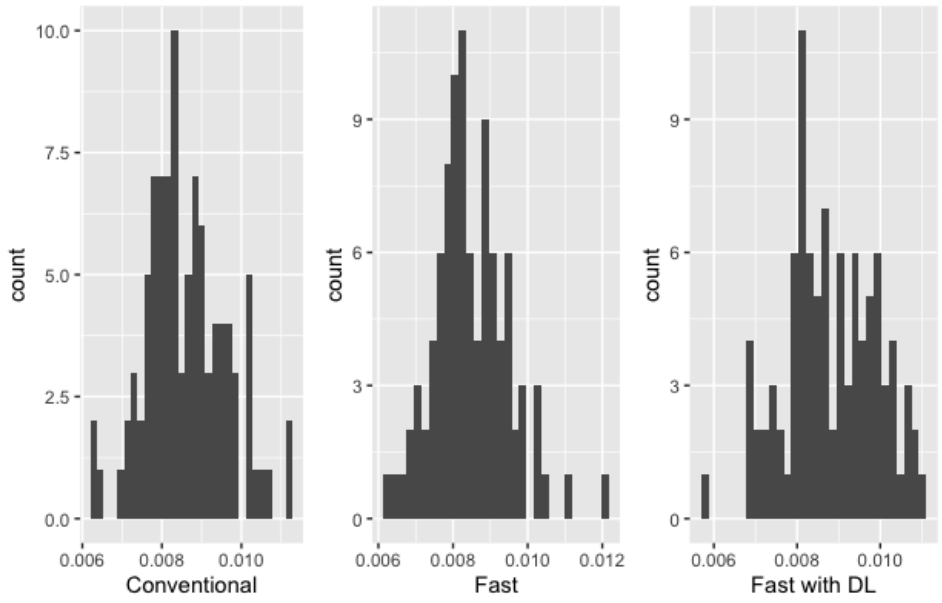
	Conventional N=96	**Fast N=96**	**Fast with DL N=96**
Normalized Brain Volume	0.745 ± 0.036	0.738 ± 0.031	0.732 ± 0.03
Normalized Thalamic Volume	0.0086 ± 0.0010	0.0085 ± 0.0010	0.0088 ± 0.0011
Normalized Gray Matter Volume	0.415 ± 0.026	0.376 ± 0.026	0.397 ± 0.024
Normalized White Matter Volume	0.291 ± 0.027	0.323 ± 0.031	0.296 ± 0.026

Figure 1 shows the distributions of the various MRI metrics by each of the three MRI methods. Generally, most distributions were unimodal and symmetric. The outlier present in the fast scan for the normalized white matter volume measurement (Figure 2 (D)) was investigated but the other volume measurements for normalized brain volume, normalized thalamic volume,

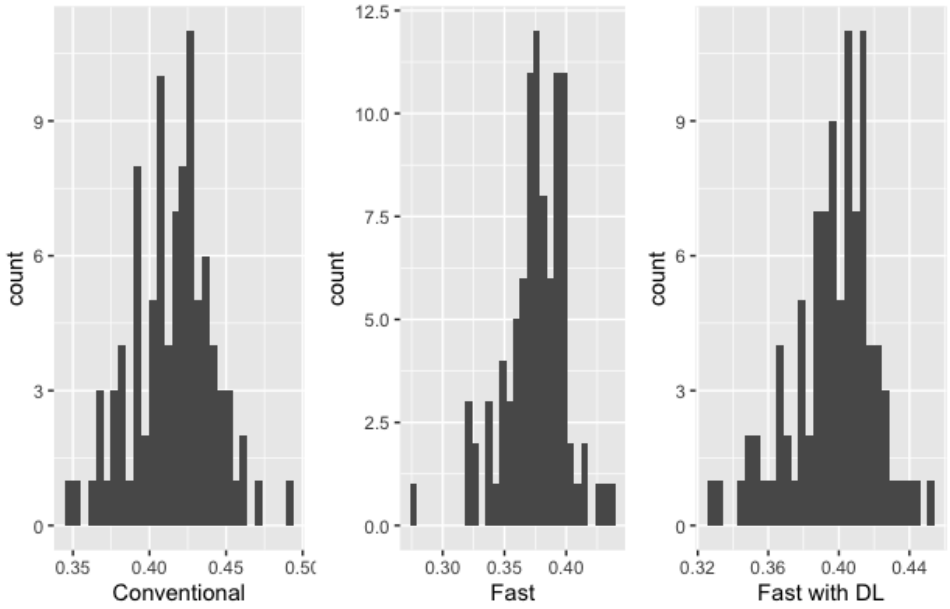
and normalized gray matter in that scan were within the standard range. Thus, we decided to keep that fast scan in the tests we performed.



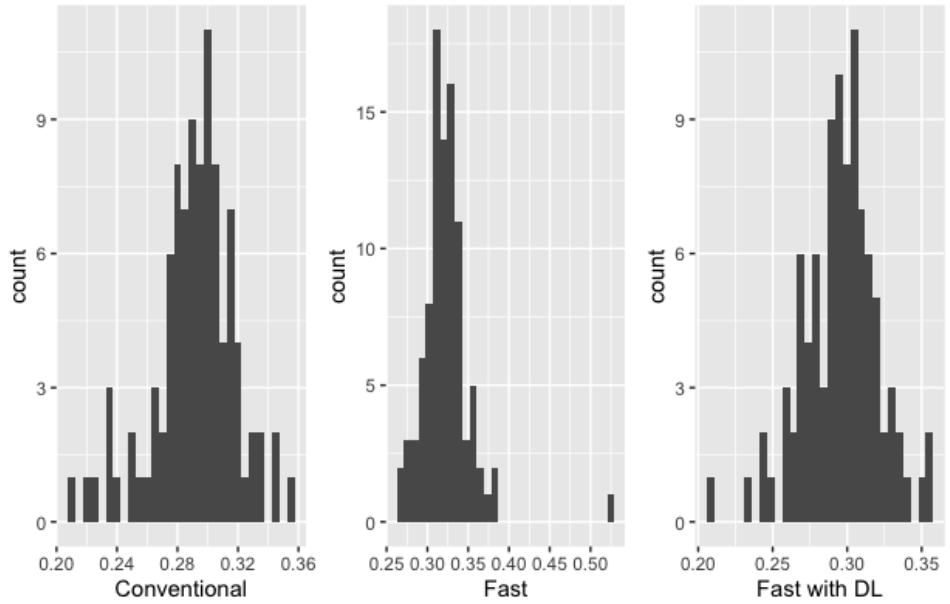
A. Normalized Brain Volume (NBV)



B. Normalized Thalamic Volume (NThV)



C. Normalized Gray Matter Volume (NGMV)



D. Normalized White Matter Volume (NWMV)

Figure 2 Distribution of Volumetric Measures by MRI metric and scan type

3.2 Comparison of MRI Methods for Volumetric Measures

3.2.1 Paired T-Tests

Figure 3 shows representative images of the three MRI acquisition techniques: conventional, fast scan, and fast scan with DL. Axial refers to the horizontal scan that occurs parallel to the ground and goes from the head to the feet. Coronal refers to the scan that occurs perpendicular to the ground and goes from the front to the back of the body. Sagittal is also perpendicular to the ground but instead scans the middle of the brain. The conventional scan took 2:57 minutes in total acquisition time, while both the fast scan and fast scan with DL took 1:13 minutes. Paired t-tests indicated a true difference in mean volumes among these methods (Table 3) . Paired t-tests comparing fast scans with conventional scans were significant for NBV, NThV, NGMV, and NWMV ($p=0.0006$, $p=0.01$, $p<0.0001$, and $p<0.0001$, respectively). Similarly, paired t-tests comparing fast scans with DL and conventional scans were significant for NBV, NThV, NGMV, and NWMV ($p<0.0001$, $p=0.002$, $p<0.0001$ and $p<0.0001$, respectively). Compared to the conventional scans, the mean percentage difference in volumes for NBV, NThV, NGMV, and NWMV for fast scans with DL were -1.789%, 2.265%, -4.273%, and 1.478%, respectively. Lastly, paired t-tests comparing fast scans with DL and fast scans were significant for NBV, NThV, NGMV, and NWMV ($p=0.0004$, $p<0.0001$, $p<0.0001$, and $p<0.0001$, respectively).

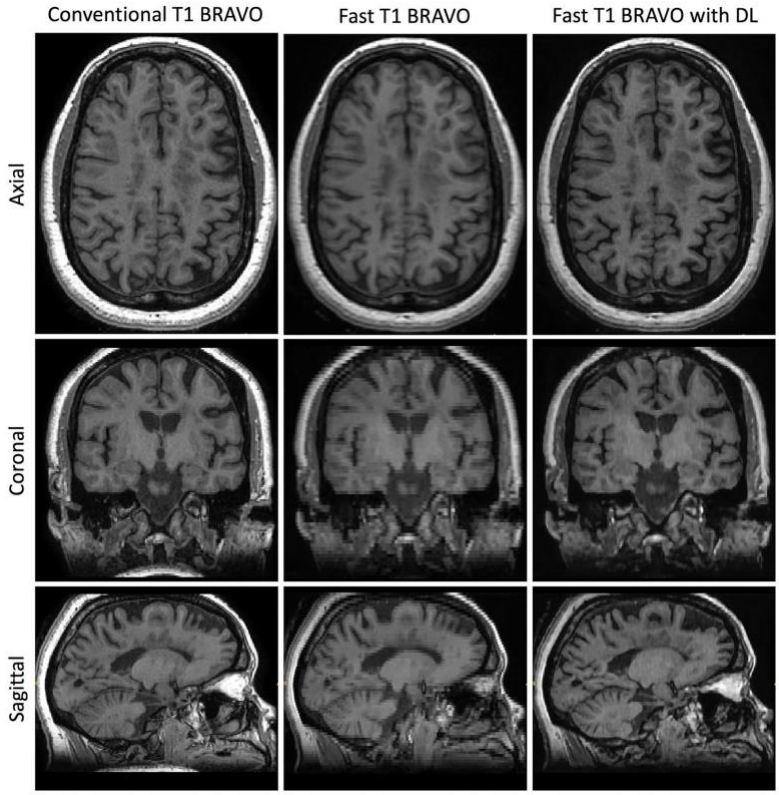


Figure 3 Representative images of MRI scan methods

Table 3 Paired Differences for Volume Measures across MRI methods

	Fast v. Conventional **Mean of the Differences ±SD** **Mean Percentage Difference**	**Fast v. Conventional** **P-values**	**Fast with DL v. Conventional** **Mean of the Differences ±SD** **Mean Percentage Difference**	**Fast with DL v. Conventional** **P-values**	**Fast with DL v. Fast** **Mean of the Differences ±SD** **Mean Percentage Difference**	**Fast with DL v. Fast** **P-values**
NBV	-0.007 ± 0.02 -0.905%	0.0006*	-0.013 ± 0.02 -1.789%	<0.0001*	-0.006 ± 0.02 -0.892%	0.0004*
NThV	-0.0002 ± 0.0006 -1.641%	0.01*	0.0002 ± 0.0006 2.265%	0.002*	0.0004 ± 0.0005 3.971%	<0.0001*
NGMV	-0.039 ± 0.01 -9.345%	<0.0001*	-0.018 ± 0.01 -4.273%	<0.0001*	0.022 ± 0.01 5.595%	<0.0001*
NWMV	0.031 ± 0.02 10.874%	<0.0001*	0.004 ± 0.01 1.478%	<0.0001*	-0.026 ± 0.02 -8.475%	<0.0001*

*Where * indicates statistical significance at* $\boldsymbol{\alpha}$ *= 0.0125*

3.2.2 Paired TOST Equivalence Tests

To determine whether fast scans with DL performed at a level comparable to that of conventional scans, we performed a series of paired two one-sided tests (TOST) equivalence tests. A 0.1 decimal fraction of the mean difference between fast and conventional scans was used as the margin of equivalence (epsilon) for the paired TOST equivalence tests. The results of the equivalence tests between fast scans with DL and conventional scans shown in Table 4 demonstrated no statistical significance for all four T1 volumetric measures: NBV, NThV, NGMV, and NWMV. Thus, with the margins specified, fast scans with DL and conventional scans provided statistically nonequivalent results for all four volumetric measures. For NBV and NGMV, fast scans with DL generally had lower volumes than conventional scans, while for NThV and NWMV, fast scans with DL had higher volumes than conventional scans.

Table 4 Paired TOST Equivalence Tests for Fast with DL vs. Conventional MRI

	Means of Differences Between Fast with DL and Conventional	TOST Lower Bounds (95%)	TOST Upper Bounds (95%)	Mean Difference Between Fast and Conventional	Epsilon (Decimal Fraction of Mean Difference Between Fast and Conventional)	P-values for equivalence between Fast with DL and Conventional
NBV	-0.013	-0.016	-0.010	0.007	0.0007 (.1)	1.000
NThV	0.0002	0.00009	0.0003	0.0002	0.0002 (.1)	0.998
NGMV	-0.018	-0.020	-0.016	0.039	0.004 (.1)	1.000
NWMV	0.004	0.003	0.006	0.031	0.003 (.1)	0.882

3.3 MRI-Clinical Correlations Comparisons

3.3.1 Repeated Measures Multivariate Linear Mixed-Effects Model - Patient Disease Determined Steps

We investigated the association between the four MRI metrics and the clinically relevant patient-reported outcomes. We examined the association with physical and gait impairment using the clinical measure of Patient Disease Determined Steps (PDDS). Age, sex, race/ethnicity, disease duration in years, RMS clinical type status, standard-efficacy treatment status, and high-efficacy treatment status were used as covariates in the model. Interaction terms were included between the clinical measure (PDDS) and the type of MRI scan as a fixed effect in our model, with conventional scan as the reference group. Table 5 shows the estimates of the fixed effects and p-values for each of the four linear mixed-effects models.

PDDS was a statistically significant predictor in all models indicating that each one-point increase in PDDS is associated with lower volumetric measures for each of the four MRI metrics. For NBV, standard-efficacy treatment status was a statistically significant predictor when controlling for the other variables in the model (estimate=-0.03, p=0.003). For NGMV, age at first scan was a statistically significant predictor of volume (estimate=-0.0008, p=0.0003). For NWMV, standard-efficacy treatment status was a statistically significant predictor of volume (estimate= -0.02, p=0.003). The interaction term between fast scans and conventional scans was statistically significant for NBV, indicating a difference in the association between NBV with PDDS when comparing fast scans to conventional scans. The interaction terms between fast scans with DL and conventional scans were not statistically significant for any of the four MRI metrics indicating no

difference in the association between volumetric measures and PDDS when comparing fast scans with DL to conventional scans.

Table 5 Linear Mixed-Effects Model with Conventional MRI scan as Baseline and Interaction with PDDS

Variable	**NBV**		**NThV**		**NGMV**		**NWMV**	
	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value
Intercept	0.8 (.02)	<0.0001*	0.01 (.0006)	<0.0001*	0.5 (.01)	<0.0001*	0.3 (.02)	<0.0001*
Fast	-0.01 (.002)	<0.0001*	-0.0003 (.00008)	0.005*	-0.04 (.001)	<0.0001*	0.03 (.002)	<0.0001*
Fast with DL	-0.01 (.002)	<0.0001*	0.0002 (.00008)	0.007*	-0.02 (.001)	<0.0001*	0.003 (.002)	0.079
PDDS	-0.007 (.002)	0.0006*	-0.0002 (.00006)	0.004*	-0.004 (.001)	0.004*	-0.005 (.002)	0.007*
Age (at first scan)	-0.0003 (.0003)	0.279	-0.00002 (.000009)	0.035	-0.0008 (.0002)	0.0003*	0.0003 (.0003)	0.289
Standard-Efficacy Treatment	-0.03 (.008)	0.003*	-0.0006 (.0002)	0.016	-0.004 (.006)	0.489	-0.02 (.007)	0.003*
Fast x PDDS	0.003 (.001)	0.002*	0.00007 (.00003)	0.025	0.001 (.0006)	0.111	0.002 (.0008)	0.019
Fast with DL x PDDS	0.002 (.001)	0.098	0.000008 (.00003)	0.790	0.001 (.0006)	0.040	0.0006 (.0008)	0.460

*Where x indicates that there is an interaction between the two variables and * indicates statistical significance at* $\boldsymbol{\alpha}$ *= 0.0125*

3.3.2 Repeated Measures Multivariate Linear Mixed-Effects Model - Multiple Sclerosis Rating Scale-Revised

We also examined the association of the volume measures with the MS symptom burden using the clinically relevant patient-reported outcome of MSRS-R. Age, sex, race/ethnicity, disease duration in years, RMS clinical type status, standard-efficacy treatment status, and high-efficacy treatment status were used as covariates in the model. Interaction terms were included between the clinical measure (MSRS-R) and the type of MRI scan as a fixed effect in our model,

with conventional scan as the reference group. Table 6 shows the estimates of the fixed effects and p-values for each of the four linear mixed-effects models. For NBV, standard-efficacy treatment status was a statistically significant predictor when controlling for the other variables in the model (estimate=-0.03, p=0.002). For NThV, age at first scan (estimate=-0.00003, p=0.003) and standard-efficacy treatment status (estimate=-0.0007, p=0.011) were statistically significant predictors of volume. For NGMV, age at first scan was a statistically significant predictor of volume (estimate=-0.0009, p<0.0001). For NWMV, standard-efficacy treatment status was a statistically significant predictor of volume (estimate=-0.02, p=0.002). The interaction terms between fast scans and conventional scans were not statistically significant for any of the four MRI metrics indicating no statistically significant difference in the association between volumetric measures and MSRS-R when comparing fast scans to conventional scans. The interaction terms between fast scans with DL and conventional scans were not statistically significant for any of the four MRI metrics indicating no statistically significant difference in the association between volumetric measures and MSRS-R when comparing fast scans with DL to conventional scans.

Table 6 Linear Mixed-Effects Model with Conventional MRI scan as Baseline and Interaction with MSRS-R

Variable	**NBV**		**NThV**		**NGMV**		**NWMV**	
	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value
Intercept	0.8 (.02)	<0.0001*	0.01 (.0006)	<0.0001*	0.5 (.01)	<0.0001*	0.3 (.02)	<0.0001*
Fast	-0.01 (.003)	<0.0001*	-0.0003 (.00008)	0.0006*	-0.04 (.002)	<0.0001*	0.03 (.002)	<0.0001*
Fast with DL	-0.01 (.003)	<0.0001*	0.0001 (.00008)	0.247	-0.02 (.002)	<0.0001*	0.004 (.002)	0.110
MSRS-R	-0.001 (.0006)	0.032	-0.00004 (.00002)	0.044	-0.0009 (.0004)	0.057	-0.0008 (.0006)	0.142
Age (at first scan)	-0.0005 (.0003)	0.055	-0.00003 (.000008)	0.003*	-0.0009 (.0002)	<0.0001*	0.00009 (.0002)	0.710
Sex	0.002 (.007)	0.827	0.00003 (.0002)	0.907	0.006 (.005)	0.254	-0.006 (.007)	0.396

Non-Hispanic white	0.005 (.005)	0.414	0.0001 (.0002)	0.473	0.001 (.004)	0.766	0.003 (.005)	0.576
Disease Duration	-0.0004 (.0002)	0.063	-0.000005 (.000006)	0.408	-0.00007 (.0001)	0.630	-0.0003 (.0002)	0.077
Clinical Type RMS	0.008 (.01)	0.494	0.00002 (.0004)	0.966	0.006 (.008)	0.445	0.005 (.01)	0.614
Standard-Efficacy Treatment	-0.03 (.008)	0.002*	-0.0007 (.0003)	0.011*	-0.005 (.006)	0.381	-0.02 (.007)	0.002*
High-Efficacy Treatment	-0.02 (.01)	0.052	-0.0008 (.0003)	0.009*	-0.007 (.007)	0.321	-0.02 (.009)	0.072
Fast x MSRS-R	0.0007 (.0003)	0.023	0.00002 (.00001)	0.026	0.0003 (.0002)	0.103	0.0004 (.0003)	0.183
Fast with DL x MSRS-R	0.0003 (.0003)	0.298	0.00002 (.00001)	0.071	0.0002 (.0002)	0.303	0.0001 (.0003)	0.618

*Where x indicates that there is an interaction between the two variables and * indicates statistical significance at* $\boldsymbol{\alpha} = 0.0125$

3.3.3 Multivariate TOST Equivalence Tests on Interaction Terms

To test the equivalence in associations between fast scans with DL and conventional scans with the clinical measures PDDS and MSRS-R, we performed a series of TOST equivalence tests on the interaction terms between the fast with DL volumetric measures and the clinical measures. Table 7 shows the results of these TOST equivalence tests. All equivalence tests were not statistically significant, indicating that the associations between fast scan with DL and conventional scan volumes with the clinical measures tested were not equivalent.

Table 7 Multivariate Equivalence Tests for Conventional Scans and Fast Scans with DL

	Measure Tested	**Estimated Difference between Fast and Conventional (SE)**	**Estimated Difference between Fast with DL and Conventional (SE)**	**P-values for Equivalence between Fast with DL and Conventional using fraction of .1**
NBV	PDDS	0.0031 (.001)	0.0016 (.001)	0.080
	MSRS-R	0.0007 (.0003)	0.0003 (.0003)	0.197
NThV	PDDS	0.00007 (.00003)	0.000008 (.00003)	0.486

	MSRS-R	0.00002 (.00001)	0.00002 (.00001)	0.048
NGMV	PDDS	0.0010 (.0006)	0.0013 (.0006)	0.023
	MSRS-R	0.0003 (.0002)	0.0002 (.0002)	0.181
NWMV	PDDS	0.0020 (.0008)	0.0006 (.0008)	0.302
	MSRS-R	0.0004 (.0003)	0.0001 (.0003)	0.351

4.0 Discussion

This analysis aimed to assess the equivalence of quality in images between conventional MRI scans and MRI scans enhanced with deep learning (DL) in a cohort of multiple sclerosis patients and explore the equivalence of quality in images between conventional MRI scans and MRI scans enhanced with deep learning (DL) in a cohort of multiple sclerosis patients. We found that an accelerated MRI scan utilizing a deep learning model did not preserve the image quality observed in a conventional MRI scan. The correlation between the volumetric metrics and clinically relevant outcomes in people with multiple sclerosis was preserved in fast scans with DL. However, we found that the correlation between the volumetric metrics and clinically relevant outcomes were not equivalent. To our knowledge, this is the first report of deep learning application to improve the image quality of accelerated scans in clinical brain MRIs for people with multiple sclerosis.

There are direct clinical implications of the acceleration of MRI scans, particularly for conditions such as MS where disease monitoring using MRI scans is the standard of care. Methods such as compressed sensing and parallel imaging methods aim to reconstruct higher quality images from smaller amounts of raw MRI data[30-32]. There is still a concern for clinical practicality due to long reconstruction times and poor image quality, however[33]. Deep learning methods started to address these issues by incorporating different types of CNN structure[3-6]. These methods reduce scan time by under-sampling k-space in raw MRI data and then they reconstruct a higher-quality image using novel DL models. Unlike feed-forward approaches, deep back-projection network utilizes error feedback to self-correct at multiple layers of the neural network. It helps to improve image quality through sharpness-enhancement of legion volume and noise elimination[11]. To our

knowledge, this study is a novel application of deep back-projection network to clinical brain MRIs for MS.

Despite the differences in volumetric metrics between fast scans with DL and conventional scans, the correlations between the key volumetrics and clinical outcomes did not differ according to the repeated measures linear mixed-effects models. This suggests that there was either an offset in the values caused by the different scan types or the calculated volumetric differences were too small to impact the MRI-clinical outcome correlations. Though the coefficients from the mixed-effects models proved that there was no difference in correlations, we found that the correlations between fast scans with DL and conventional scans with the clinical outcomes were not equivalent. There have not been previous studies that have explored the correlation between these clinical outcomes with brain volume measurements. A few covariates were statistically significant in the model, specifically: age at first scan, standard-efficacy treatment status, and high-efficacy treatment status. There are no previous studies that looked at the effect of treatment type on brain volume measurements. One study looked at the changes in brain volume over time in patients with MS and found that age, sex, frequency of ongoing inflammation, multiple sclerosis clinical type, and randomized treatment assignment did not have any effect on brain volume[34]. This is consistent with our findings in regards to sex and multiple sclerosis clinical type in the linear mixed-effects models.

There were limitations to this study. The study had a modest sample size, limiting the power of some of the statistical analyses. This is particularly evident in the equivalence tests, where the sample size needed to achieve a particular power level is dependent on the equivalence margin[35]. The apparent disagreement between the traditional test and equivalence test for the MRI-clinical outcome correlations may be due to this lack of adequate sample size. The study population

is quite representative of the population with MS, so our conclusions are not limited by that in any way. Second, we did not have any prior clinical information to help inform the equivalence margins chosen for the equivalence tests.

In conclusion, we demonstrated the clinical application of a deep learning model utilizing deep back-projection networks to restore the image quality of accelerated MRI scans for MS. Shortened MRI scans improve patient comfort and satisfaction while reducing issues introduced by involuntary motion that often occurs in the latter portion of a prolonged MRI study. Utilizing shortened scans would enable efficient utilization of MRI resources by reducing unnecessary wait time and improving access to clinical imaging for diagnostic and monitoring purposes, not only for people with MS but also for other patient populations. However, we found that fast scans with DL do not produce images of equivalent quality to conventional scans. We also found that the correlations between the key volumetrics and clinical outcomes did not differ between fast scans with DL and conventional scans but there was not sufficient evidence to prove that the correlations are equivalent. We anticipate future studies that test the ability of MRI scans utilizing DL models to replace conventional MRI scans[36,37].

Appendix A R Code

```
```{r setup, include=FALSE}
library(stringr)
library(readxl)
library(tidyverse)
library(skimr)
library(knitr)
library(equivalence)
library(pastecs)
MRIdata <- read_excel("Batch1-4_20200420.xlsx",sheet=1)
i <- c(8:48)
MRIdata[ , i] <- apply(MRIdata[ , i], 2,            # Specify
own function within apply
                  function(x) as.numeric(as.character(x)))
MRIdata<-MRIdata[-c(97:124),-c(11:36)]
MRIUniq <- MRIdata[!duplicated(MRIdata$`Patient ID`), ]
clin <- read_excel("Subtle_metadata_updated.xls",sheet=1)
clinUniq <- clin[!duplicated(clin$`Patient ID`), ]
clinical<-merge(x =MRIUniq, y = clinUniq, by="Patient ID")
clinical$amb<-ifelse(clinical$`PDDS Median`>=4, 1, 0)
clinical$higheff<-
ifelse(clinical$Treatment==(c(13,15,11,16,17)), 1, 0)
clinical$cau<-ifelse(clinical$Race=="Caucasian", 1, 0)
clinical$nonhis<-ifelse(clinical$Ethnicity=="Non-
Hispanic"|clinical$Ethnicity=="Non-hispanic", 1, 0)
clinical$nonhiswhite<-ifelse((clinical$Ethnicity=="Non-Hispanic"
| clinical$Ethnicity=="Non-
hispanic")&(clinical$Race=="Caucasian"), 1, 0)
clinical$types <- ifelse(clinical[,35]==1,1,0)
clinical$notreat <- ifelse(clinical$Treatment==99,1,0)
clinicalamb <- clinical[which(clinical$amb==1), ]
clinicalnoamb <- clinical[which(clinical$amb==0), ]
```

```{r}
options(scipen=999)
histogram(clinical$`PDDS Median`)
summarytab<-stat.desc(clinical, basic=F)
summary(clinical$`PDDS Median`)
summary(clinicalnoamb$`PDDS Median`)
summary(clinicalamb$`PDDS Median`)
sd(clinical$`PDDS Median`)
sd(clinicalnoamb$`PDDS Median`)
```
```

```
sd(clinicalamb$`PDDS Median`)
summary(clinical$`Age (at first scan)`)
summary(clinicalnoamb$`Age (at first scan)`)
summary(clinicalamb$`Age (at first scan)`)
sd(clinical$`Age (at first scan)`)
sd(clinicalnoamb$`Age (at first scan)`)
sd(clinicalamb$`Age (at first scan)`)
summary(clinical$`MSRS-R Median`)
summary(clinicalnoamb$`MSRS-R Median`)
summary(clinicalamb$`MSRS-R Median`)
summary(clinical$`Disease Duration (date of MRI - date of first
symptoms), years`)
summary(clinicalnoamb$`Disease Duration (date of MRI - date of
first symptoms), years`)
summary(clinicalamb$`Disease Duration (date of MRI - date of
first symptoms), years`)
sd(clinical$`Disease Duration (date of MRI - date of first
symptoms), years`)
sd(clinicalnoamb$`Disease Duration (date of MRI - date of first
symptoms), years`)
sd(clinicalamb$`Disease Duration (date of MRI - date of first
symptoms), years`)
summarytabamb<-stat.desc(clinicalamb,basic=F)
summarytabnoamb<-stat.desc(clinicalnoamb,basic=F)
table(clinical$Sex)
table(clinicalnoamb$Sex)
table(clinicalamb$Sex)
prop.table(table(clinical$Sex))
prop.table(table(clinicalnoamb$Sex))
prop.table(table(clinicalamb$Sex))
table(clinical$Race)
table(clinicalnoamb$Race)
table(clinicalamb$Race)
prop.table(table(clinical$Race))
prop.table(table(clinicalnoamb$Race))
prop.table(table(clinicalamb$Race))
table(clinical$nonhis)
table(clinicalnoamb$nonhis)
table(clinicalamb$nonhis)
prop.table(table(clinical$nonhis))
prop.table(table(clinicalnoamb$nonhis))
prop.table(table(clinicalamb$nonhis))
table(clinical$types)
table(clinicalnoamb$types)
table(clinicalamb$types)
prop.table(table(clinical$types))
prop.table(table(clinicalnoamb$types))
```

```
prop.table(table(clinicalamb$types))
table(clinical$notreat)
table(clinicalnoamb$notreat)
table(clinicalamb$notreat)
prop.table(table(clinical$notreat))
prop.table(table(clinicalnoamb$notreat))
prop.table(table(clinicalamb$notreat))
table(clinical$higheff)
table(clinicalnoamb$higheff)
table(clinicalamb$higheff)
prop.table(table(clinical$higheff))
prop.table(table(clinicalnoamb$higheff))
prop.table(table(clinicalamb$higheff))
table(clinical$nonhiswhite)
table(clinicalnoamb$nonhiswhite)
table(clinicalamb$nonhiswhite)
prop.table(table(clinical$nonhiswhite))
prop.table(table(clinicalnoamb$nonhiswhite))
prop.table(table(clinicalamb$nonhiswhite))
```
```{r}
mean(MRIdata$T2LV_Conv, na.rm = TRUE)
sd(MRIdata$T2LV_Conv, na.rm = TRUE)

mean(MRIdata$NBV_Conv, na.rm = TRUE)
sd(MRIdata$NBV_Conv, na.rm = TRUE)

mean(MRIdata$NBV_Subtle, na.rm = TRUE)
sd(MRIdata$NBV_Subtle, na.rm = TRUE)
mean(MRIdata$NBV_SubtleMR, na.rm = TRUE)
sd(MRIdata$NBV_SubtleMR, na.rm = TRUE)

mean(MRIdata$NThV_Conv, na.rm = TRUE)
sd(MRIdata$NThV_Conv, na.rm = TRUE)
mean(MRIdata$NThV_Subtle, na.rm = TRUE)
sd(MRIdata$NThV_Subtle, na.rm = TRUE)
mean(MRIdata$NThV_SubtleMR, na.rm = TRUE)
sd(MRIdata$NThV_SubtleMR, na.rm = TRUE)

mean(MRIdata$NGMV_Conv, na.rm = TRUE)
sd(MRIdata$NGMV_Conv, na.rm = TRUE)
mean(MRIdata$NGMV_Subtle, na.rm = TRUE)
sd(MRIdata$NGMV_Subtle, na.rm = TRUE)
mean(MRIdata$NGMV_SubtleMR, na.rm = TRUE)
sd(MRIdata$NGMV_SubtleMR, na.rm = TRUE)

mean(MRIdata$NWMV_Conv, na.rm = TRUE)
```

```
sd(MRIdata$NWMV_Conv, na.rm = TRUE)
mean(MRIdata$NWMV_Subtle, na.rm = TRUE)
sd(MRIdata$NWMV_Subtle, na.rm = TRUE)
mean(MRIdata$NWMV_SubtleMR, na.rm = TRUE)
sd(MRIdata$NWMV_SubtleMR, na.rm = TRUE)

((mean(MRIdata$T2LV_Subtle, na.rm = TRUE) -
mean(MRIdata$T2LV_Conv, na.rm = TRUE))/ mean(MRIdata$T2LV_Conv,
na.rm = TRUE)) * 100
(mean(MRIdata$T2LV_Subtle, na.rm = TRUE) -
mean(MRIdata$T2LV_Conv, na.rm = TRUE))

((mean(MRIdata$NBV_Subtle, na.rm = TRUE) -
mean(MRIdata$NBV_Conv, na.rm = TRUE))/ mean(MRIdata$NBV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NBV_Subtle, na.rm = TRUE) - mean(MRIdata$NBV_Conv,
na.rm = TRUE)
((mean(MRIdata$NThV_Subtle, na.rm = TRUE) -
mean(MRIdata$NThV_Conv, na.rm = TRUE))/mean(MRIdata$NThV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NThV_Subtle, na.rm = TRUE) -
mean(MRIdata$NThV_Conv, na.rm = TRUE)
((mean(MRIdata$NGMV_Subtle, na.rm = TRUE) -
mean(MRIdata$NGMV_Conv, na.rm = TRUE))/mean(MRIdata$NGMV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NGMV_Subtle, na.rm = TRUE) -
mean(MRIdata$NGMV_Conv, na.rm = TRUE)
((mean(MRIdata$NWMV_Subtle, na.rm = TRUE) -
mean(MRIdata$NWMV_Conv, na.rm = TRUE))/mean(MRIdata$NWMV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NWMV_Subtle, na.rm = TRUE) -
mean(MRIdata$NWMV_Conv, na.rm = TRUE)
# (Fast - conventional)/conventional * 100

((mean(MRIdata$T2LV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$T2LV_Conv, na.rm = TRUE))/ mean(MRIdata$T2LV_Conv,
na.rm = TRUE)) * 100
(mean(MRIdata$T2LV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$T2LV_Conv, na.rm = TRUE))

((mean(MRIdata$NBV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NBV_Conv, na.rm = TRUE))/ mean(MRIdata$NBV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NBV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NBV_Conv, na.rm = TRUE)
```

```
((mean(MRIdata$NThV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NThV_Conv, na.rm = TRUE))/mean(MRIdata$NThV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NThV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NThV_Conv, na.rm = TRUE)

((mean(MRIdata$NGMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NGMV_Conv, na.rm = TRUE))/mean(MRIdata$NGMV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NGMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NGMV_Conv, na.rm = TRUE)

((mean(MRIdata$NWMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NWMV_Conv, na.rm = TRUE))/mean(MRIdata$NWMV_Conv,
na.rm = TRUE)) * 100
mean(MRIdata$NWMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NWMV_Conv, na.rm = TRUE)

((mean(MRIdata$T2LV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$T2LV_Subtle, na.rm = TRUE))/
mean(MRIdata$T2LV_Subtle, na.rm = TRUE)) * 100
(mean(MRIdata$T2LV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$T2LV_Subtle, na.rm = TRUE))

((mean(MRIdata$NBV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NBV_Subtle, na.rm = TRUE))/
mean(MRIdata$NBV_Subtle, na.rm = TRUE)) * 100
abs(mean(MRIdata$NBV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NBV_Subtle, na.rm = TRUE))

((mean(MRIdata$NThV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NThV_Subtle, na.rm =
TRUE))/mean(MRIdata$NThV_Subtle, na.rm = TRUE)) * 100
mean(MRIdata$NThV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NThV_Subtle, na.rm = TRUE)

((mean(MRIdata$NGMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NGMV_Subtle, na.rm =
TRUE))/mean(MRIdata$NGMV_Subtle, na.rm = TRUE)) * 100
mean(MRIdata$NGMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NGMV_Subtle, na.rm = TRUE)

((mean(MRIdata$NWMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NWMV_Subtle, na.rm =
TRUE))/mean(MRIdata$NWMV_Subtle, na.rm = TRUE)) * 100
mean(MRIdata$NWMV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NWMV_Subtle, na.rm = TRUE)
```

```
```{r}
median(MRIdata$NBV_Conv, na.rm = TRUE)
IQR(MRIdata$NBV_Conv, na.rm = TRUE)
median(MRIdata$NBV_Subtle, na.rm = TRUE)
IQR(MRIdata$NBV_Subtle, na.rm = TRUE)
median(MRIdata$NBV_SubtleMR, na.rm = TRUE)
IQR(MRIdata$NBV_SubtleMR, na.rm = TRUE)
median(MRIdata$NThV_Conv, na.rm = TRUE)
IQR(MRIdata$NThV_Conv, na.rm = TRUE)
median(MRIdata$NThV_Subtle, na.rm = TRUE)
IQR(MRIdata$NThV_Subtle, na.rm = TRUE)
median(MRIdata$NThV_SubtleMR, na.rm = TRUE)
IQR(MRIdata$NThV_SubtleMR, na.rm = TRUE)
median(MRIdata$NGMV_Conv, na.rm = TRUE)
IQR(MRIdata$NGMV_Conv, na.rm = TRUE)
median(MRIdata$NGMV_Subtle, na.rm = TRUE)
IQR(MRIdata$NGMV_Subtle, na.rm = TRUE)
median(MRIdata$NGMV_SubtleMR, na.rm = TRUE)
IQR(MRIdata$NGMV_SubtleMR, na.rm = TRUE)
median(MRIdata$NWMV_Conv, na.rm = TRUE)
IQR(MRIdata$NWMV_Conv, na.rm = TRUE)
median(MRIdata$NWMV_Subtle, na.rm = TRUE)
IQR(MRIdata$NWMV_Subtle, na.rm = TRUE)
median(MRIdata$NWMV_SubtleMR, na.rm = TRUE)
IQR(MRIdata$NWMV_SubtleMR, na.rm = TRUE)
```
Paired t-tests for NBV
```{r}
t.test(MRIdata$NBV_Subtle,MRIdata$NBV_Conv, paired=TRUE)
mean(MRIdata$NBV_Subtle-MRIdata$NBV_Conv,na.rm=TRUE)
(mean(MRIdata$NBV_Subtle, na.rm = TRUE) - mean(MRIdata$NBV_Conv,
na.rm = TRUE))
sd(MRIdata$NBV_Subtle-MRIdata$NBV_Conv,na.rm=TRUE)
t.test(MRIdata$NBV_SubtleMR,MRIdata$NBV_Conv, paired=TRUE)
(mean(MRIdata$NBV_SubtleMR, na.rm = TRUE) -
mean(MRIdata$NBV_Conv, na.rm = TRUE))
mean(MRIdata$NBV_SubtleMR-MRIdata$NBV_Conv,na.rm=TRUE)
sd(MRIdata$NBV_SubtleMR-MRIdata$NBV_Conv,na.rm=TRUE)
t.test(MRIdata$NBV_SubtleMR,MRIdata$NBV_Subtle, paired=TRUE)
mean(MRIdata$NBV_SubtleMR-MRIdata$NBV_Subtle,na.rm=TRUE)
sd(MRIdata$NBV_SubtleMR-MRIdata$NBV_Subtle,na.rm=TRUE)
```
NThV
```{r}
t.test(MRIdata$NThV_Subtle,MRIdata$NThV_Conv, paired=TRUE)
mean(MRIdata$NThV_Subtle-MRIdata$NThV_Conv,na.rm=TRUE)
```
```

```
sd(MRIdata$NThV_Subtle-MRIdata$NThV_Conv,na.rm=TRUE)
t.test(MRIdata$NThV_SubtleMR,MRIdata$NThV_Conv, paired=TRUE)
mean(MRIdata$NThV_SubtleMR-MRIdata$NThV_Conv,na.rm=TRUE)
sd(MRIdata$NThV_SubtleMR-MRIdata$NThV_Conv,na.rm=TRUE)
t.test(MRIdata$NThV_SubtleMR,MRIdata$NThV_Subtle, paired=TRUE)
mean(MRIdata$NThV_SubtleMR-MRIdata$NThV_Subtle,na.rm=TRUE)
sd(MRIdata$NThV_SubtleMR-MRIdata$NThV_Subtle,na.rm=TRUE)
```

NGMV

```{r}
t.test(MRIdata$NGMV_Subtle,MRIdata$NGMV_Conv, paired=TRUE)
mean(MRIdata$NGMV_Subtle-MRIdata$NGMV_Conv,na.rm=TRUE)
sd(MRIdata$NGMV_Subtle-MRIdata$NGMV_Conv,na.rm=TRUE)
t.test(MRIdata$NGMV_SubtleMR,MRIdata$NGMV_Conv, paired=TRUE)
mean(MRIdata$NGMV_SubtleMR-MRIdata$NGMV_Conv,na.rm=TRUE)
sd(MRIdata$NGMV_SubtleMR-MRIdata$NGMV_Conv,na.rm=TRUE)
t.test(MRIdata$NGMV_SubtleMR,MRIdata$NGMV_Subtle, paired=TRUE)
mean(MRIdata$NGMV_SubtleMR-MRIdata$NGMV_Subtle,na.rm=TRUE)
sd(MRIdata$NGMV_SubtleMR-MRIdata$NGMV_Subtle,na.rm=TRUE)
```

NWMV

```{r}
t.test(MRIdata$NWMV_Subtle,MRIdata$NWMV_Conv, paired=TRUE)
mean(MRIdata$NWMV_Subtle-MRIdata$NWMV_Conv,na.rm=TRUE)
sd(MRIdata$NWMV_Subtle-MRIdata$NWMV_Conv,na.rm=TRUE)
t.test(MRIdata$NWMV_SubtleMR,MRIdata$NWMV_Conv, paired=TRUE)
mean(MRIdata$NWMV_SubtleMR-MRIdata$NWMV_Conv,na.rm=TRUE)
sd(MRIdata$NWMV_SubtleMR-MRIdata$NWMV_Conv,na.rm=TRUE)
t.test(MRIdata$NWMV_SubtleMR,MRIdata$NWMV_Subtle, paired=TRUE)
mean(MRIdata$NWMV_SubtleMR-MRIdata$NWMV_Subtle,na.rm=TRUE)
sd(MRIdata$NWMV_SubtleMR-MRIdata$NWMV_Subtle,na.rm=TRUE)
```

```{r}
library(ggpubr)
p1 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NBV_Conv)) + scale_x_continuous(name =
"Conventional")
p2 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NBV_Subtle)) + scale_x_continuous(name =
"Fast")
p3 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NBV_SubtleMR)) + scale_x_continuous(name
= "Fast with DL")
ggarrange(p1,p2,p3, nrow = 1)
```

```{r}
p1 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NThV_Conv)) + scale_x_continuous(breaks =
c(0.006, 0.008, 0.010, 0.012), name = "Conventional")
p2 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NThV_Subtle)) + scale_x_continuous(name =
"Fast")
p3 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NThV_SubtleMR)) +
scale_x_continuous(breaks = c(0.006, 0.008, 0.010, 0.012), name
= "Fast with DL")
ggarrange(p1,p2,p3, nrow = 1)
```

```{r}
p1 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NGMV_Conv)) + scale_x_continuous(name =
"Conventional")
p2 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NGMV_Subtle)) + scale_x_continuous(name =
"Fast")
p3 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NGMV_SubtleMR)) + scale_x_continuous(name
= "Fast with DL")
ggarrange(p1,p2,p3, nrow =1 )
```

```{r}
p1 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NWMV_Conv)) + scale_x_continuous(name =
"Conventional")
p2 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NWMV_Subtle)) + scale_x_continuous(name =
"Fast")
p3 = ggplot(data=MRIdata) +
  geom_histogram(aes(x=NWMV_SubtleMR)) + scale_x_continuous(name
= "Fast with DL")
ggarrange(p1,p2,p3, nrow=1)
```

```{r,include=FALSE,cache=TRUE}
library(readxl)
library(tidyverse)
library(skimr)
library(knitr)
library(equivalence)
```

```
MRIdata <- read_excel("Batch1-4_20200420.xlsx",sheet=1)
i <- c(8:48)
MRIdata[ , i] <- apply(MRIdata[ , i], 2,            # Specify
own function within apply
                    function(x) as.numeric(as.character(x)))
MRIdata<-MRIdata[,-c(11,15)]
```

NBV Equivalence tests - Conv vs SubtleMR {#Sig9md}

```{r}
md = abs(mean(MRIdata$NBV_Subtle - MRIdata$NBV_Conv, na.rm =
TRUE))
md
tost(MRIdata$NBV_SubtleMR, MRIdata$NBV_Conv, epsilon = .1*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NBV_SubtleMR, MRIdata$NBV_Conv,  epsilon = .2*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NBV_SubtleMR, MRIdata$NBV_Conv,epsilon = .3*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NBV_SubtleMR, MRIdata$NBV_Conv, epsilon = .4*md,
paired = T, var.equal = F, conf.level = 0.95)
```

NThV Equivalence tests - Conv vs SubtleMR {#Sig10md}

```{r}
md = abs(mean(MRIdata$NThV_Subtle - MRIdata$NThV_Conv, na.rm =
TRUE))
md
tost(MRIdata$NThV_SubtleMR, MRIdata$NThV_Conv, epsilon = .1*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NThV_SubtleMR, MRIdata$NThV_Conv,  epsilon = .2*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NThV_SubtleMR, MRIdata$NThV_Conv,epsilon = .3*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NThV_SubtleMR, MRIdata$NThV_Conv, epsilon = .4*md,
paired = T, var.equal = F, conf.level = 0.95)
```

NGMV Equivalence tests - Conv vs SubtleMR {#Sig11md}

```{r}
md = abs(mean(MRIdata$NGMV_Subtle - MRIdata$NGMV_Conv, na.rm =
TRUE))
md
.1*md
```

```
tost(MRIdata$NGMV_SubtleMR, MRIdata$NGMV_Conv, epsilon = .1*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NGMV_SubtleMR, MRIdata$NGMV_Conv,  epsilon = .2*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NGMV_SubtleMR, MRIdata$NGMV_Conv,epsilon = .3*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NGMV_SubtleMR, MRIdata$NGMV_Conv, epsilon = .4*md,
paired = T, var.equal = F, conf.level = 0.95)
```

NWMV Equivalence tests - Conv vs SubtleMR {#Sig12md}

```{r}
md = abs(mean(MRIdata$NWMV_Subtle - MRIdata$NWMV_Conv, na.rm =
TRUE))
md

t.test(MRIdata$NWMV_Subtle, MRIdata$NWMV_Conv, paired=TRUE)
tost(MRIdata$NWMV_SubtleMR, MRIdata$NWMV_Conv, epsilon = .1*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NWMV_SubtleMR, MRIdata$NWMV_Conv,  epsilon = .2*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NWMV_SubtleMR, MRIdata$NWMV_Conv,epsilon = .3*md,
paired = T, var.equal = F, conf.level = 0.95)
tost(MRIdata$NWMV_SubtleMR, MRIdata$NWMV_Conv, epsilon = .4*md,
paired = T, var.equal = F, conf.level = 0.95)
```

```{r,include=FALSE}
library(readxl)
library(tidyverse)
library(skimr)
library(knitr)
library(tidyr)
library(lmerTest)
library(car)
library(plyr)
library(PowerTOST)
library(interactions)
MRIdata <- read_excel("Batch1-4_20200420.xlsx",sheet=1)
i <- c(8:48)
MRIdata[ , i] <- apply(MRIdata[ , i], 2,
                    function(x) as.numeric(as.character(x)))
MRIdata<-MRIdata[-c(97:124),-c(11:36)]
clin <- read_excel("Subtle_metadata_updated.xls",sheet=1)
clin$amb<-ifelse(clin$`PDDS Median`>=4, 1, 0)
```

```
clin$sexbin<-ifelse(clin$Sex=="Female", 1, 0) #1 if female, 0
male
clin$nonhiswhite<-ifelse((clin$Ethnicity=="Non-Hispanic" |
clin$Ethnicity=="Non-hispanic")&(clin$Race=="Caucasian"), 1, 0)
#1 if non-hispanic european descent, 0 other
clin<-plyr::rename(clin, c("Age (at first scan)" = "age",
"Disease Duration (date of MRI - date of first symptoms),
years"="dur"))
clin$types <- ifelse(clin[,14]==1,1,0) # 1 if RRMS and 0 other
clin$standard <-
ifelse(clin$Treatment==1|clin$Treatment==2|clin$Treatment==3|cli
n$Treatment==4|clin$Treatment==5|clin$Treatment==6|clin$Treatmen
t==18|clin$Treatment==7|clin$Treatment==8|clin$Treatment==9|clin
$Treatment==10|clin$Treatment==19|clin$Treatment==20|clin$Treatm
ent==14,1,0) # 1 if standard efficacy 0 if other
clin$high <-
ifelse(clin$Treatment==16|clin$Treatment==17|clin$Treatment==15|
clin$Treatment==13|clin$Treatment==11,1,0)# 1 if high efficacy 0
if other
clinical<-merge(x =MRIdata, y = clin, by="Patient ID")
clin_long <- gather(clinical, type, volume,
T2LV_Conv:NWMV_SubtleMR, factor_key=TRUE)
```

NBV

Linear model with PDDS (continuous) as a predictor variable

```{r,echo=FALSE}
NBVmod2 <-lmer(volume ~ type * `PDDS Median` + age+ sexbin +
nonhiswhite + dur + types + standard + high + (1 | `Patient ID`)
, data=NBV)
summary(NBVmod2)
plot_model(NBVmod2, type='diag')
plot(resid(NBVmod2), NBVmod2@frame[["volume"]])
```

```{r, echo=FALSE}
linearHypothesis(NBVmod2, c("typeNBV_SubtleMR:`PDDS Median`"))
linearHypothesis(NBVmod2, c("typeNBV_Subtle:`PDDS Median`"))
linearHypothesis(NBVmod2, c("typeNBV_SubtleMR:`PDDS Median`-
.1*typeNBV_Subtle:`PDDS Median`"),test="F")/2
linearHypothesis(NBVmod2, c("typeNBV_SubtleMR:`PDDS
Median`+.1*typeNBV_Subtle:`PDDS Median`"),test="F")/2
linearHypothesis(NBVmod2, c("typeNBV_SubtleMR:`PDDS Median`-
.1*typeNBV_Subtle:`PDDS Median`"), test="F")
pt(sqrt(1.977), 226, lower.tail=F)
```

```
linearHypothesis(NBVmod2, c("typeNBV_SubtleMR:`PDDS
Median`+.1*typeNBV_Subtle:`PDDS Median`"), test="F")
pt(-sqrt(3.5399), 226, lower.tail=T)
```

Linear model with MSRS-R as a predictor variable

```{r,echo=FALSE}
NBVmsrsmod <-lmer( volume ~ type * `MSRS-R Median` + age+ sexbin
+ nonhiswhite + dur + types+ standard+high + (1 | `Patient ID`)
, data=NBV)
summary(NBVmsrsmod)
plot_model(NBVmsrsmod, type='diag')
```

```{r, echo=FALSE}
linearHypothesis(NBVmsrsmod, c("typeNBV_SubtleMR:`MSRS-R
Median`"))
linearHypothesis(NBVmsrsmod, c("typeNBV_Subtle:`MSRS-R
Median`"))
linearHypothesis(NBVmsrsmod, c("typeNBV_SubtleMR:`MSRS-R
Median`-.1*typeNBV_Subtle:`MSRS-R Median`"))/2
linearHypothesis(NBVmsrsmod, c("typeNBV_SubtleMR:`MSRS-R
Median`+.1*typeNBV_Subtle:`MSRS-R Median`"))/2
```
NThV

Linear model with PDDS (continuous) as a predictor variable

```{r,echo=FALSE}
NThVmod2 <-lmer( volume ~ type * `PDDS Median` + age+ sexbin +
nonhiswhite + dur + types + standard+high + (1 | `Patient ID`) ,
data=NThV)
summary(NThVmod2)
plot_model(NThVmod2, type='diag')
```

```{r, echo=FALSE}
linearHypothesis(NThVmod2, c("typeNThV_SubtleMR:`PDDS Median`-
.1*typeNThV_Subtle:`PDDS Median`"))/2
linearHypothesis(NThVmod2, c("typeNThV_SubtleMR:`PDDS
Median`+.1*typeNThV_Subtle:`PDDS Median`"))/2
```

```{r}
interact_plot(NThVmod2, pred = `PDDS Median`, modx = type)
```

Linear model with MSRS-R as a predictor variable

```{r,echo=FALSE}
NThVmsrsmod <-lmer( volume ~ type * `MSRS-R Median` + age+
sexbin + nonhiswhite + dur + types+ standard+high + (1 |
`Patient ID`) , data=NThV)
summary(NThVmsrsmod)
plot_model(NThVmsrsmod, type='diag')
```

```{r, echo=FALSE}
linearHypothesis(NThVmsrsmod, c("typeNThV_SubtleMR:`MSRS-R
Median`-.1*typeNThV_Subtle:`MSRS-R Median`"))/2
linearHypothesis(NThVmsrsmod, c("typeNThV_SubtleMR:`MSRS-R
Median`+.1*typeNThV_Subtle:`MSRS-R Median`"))/2
linearHypothesis(NThVmsrsmod, c("typeNThV_SubtleMR:`MSRS-R
Median`-.1*typeNThV_Subtle:`MSRS-R Median`"),test="F")/2
linearHypothesis(NThVmsrsmod, c("typeNThV_SubtleMR:`MSRS-R
Median`-.1*typeNThV_Subtle:`MSRS-R Median`"),test="F")/2
linearHypothesis(NThVmsrsmod, c("typeNThV_SubtleMR:`MSRS-R
Median`-.1*typeNThV_Subtle:`MSRS-R Median`"), test="F")
pt(sqrt(2.78), 226, lower.tail=F)
linearHypothesis(NThVmsrsmod, c("typeNThV_SubtleMR:`MSRS-R
Median`-.1*typeNThV_Subtle:`MSRS-R Median`"), test="F")
pt(-sqrt(2.78), 226, lower.tail=T)
```

```{r}
interact_plot(NThVmsrsmod, pred = `MSRS-R Median`, modx = type)
```

NGMV

Linear model with PDDS (continuous) as a predictor variable

```{r,echo=FALSE}
NGMVmod2 <-lmer( volume ~ type * `PDDS Median` + age+ sexbin +
nonhiswhite + dur + types + standard+high + (1 | `Patient ID`) ,
data=NGMV)
summary(NGMVmod2)
plot_model(NGMVmod2, type='diag')
```

```{r}
interact_plot(NGMVmod2, pred = `PDDS Median`, modx = type)
```

```{r, echo=FALSE}
linearHypothesis(NGMVmod2, c("typeNGMV_SubtleMR:`PDDS Median`-
.1*typeNGMV_Subtle:`PDDS Median`"))/2
linearHypothesis(NGMVmod2, c("typeNGMV_SubtleMR:`PDDS
Median`+.1*typeNGMV_Subtle:`PDDS Median`"))/2
```

Linear model with MSRS-R as a predictor variable

```{r,echo=FALSE}
NGMVmsrsmod <-lmer( volume ~ type * `MSRS-R Median` + age+
sexbin + nonhiswhite + dur + types + standard+high + (1 |
`Patient ID`) , data=NGMV)
summary(NGMVmsrsmod)
plot_model(NGMVmsrsmod, type='diag')
```

```{r, echo=FALSE}
linearHypothesis(NGMVmsrsmod, c("typeNGMV_SubtleMR:`MSRS-R
Median`-.1*typeNGMV_Subtle:`MSRS-R Median`"))/2
linearHypothesis(NGMVmsrsmod, c("typeNGMV_SubtleMR:`MSRS-R
Median`+.1*typeNGMV_Subtle:`MSRS-R Median`"))/2
```
NWMV

Linear model with PDDS (continuous) as a predictor variable

```{r,echo=FALSE}
NWMVmod2 <-lmer( volume ~ type * `PDDS Median`+ age+ sexbin +
nonhiswhite + dur + types+ standard+high + (1 | `Patient ID`) ,
data=NWMV)
summary(NWMVmod2)
plot_model(NWMVmod2, type='diag')
```

```{r, echo=FALSE}
linearHypothesis(NWMVmod2, c("typeNWMV_SubtleMR:`PDDS Median`-
.1*typeNWMV_Subtle:`PDDS Median`"))/2
linearHypothesis(NWMVmod2, c("typeNWMV_SubtleMR:`PDDS
Median`+.1*typeNWMV_Subtle:`PDDS Median`"))/2
```
Linear model with MSRS-R as a predictor variable

```{r,echo=FALSE}
```

```
NWMVmsrsmod <-lmer( volume ~ type * `MSRS-R Median` + age+
sexbin + nonhiswhite + dur + types + standard+high+ (1 |
`Patient ID`) , data=NWMV)
summary(NWMVmsrsmod)
plot_model(NWMVmsrsmod, type='diag')
```

```{r, echo=FALSE}
linearHypothesis(NWMVmsrsmod, c("typeNWMV_SubtleMR:`MSRS-R
Median`-.1*typeNWMV_Subtle:`MSRS-R Median`"))/2
linearHypothesis(NWMVmsrsmod, c("typeNWMV_SubtleMR:`MSRS-R
Median`+.1*typeNWMV_Subtle:`MSRS-R Median`"))/2
```

Bibliography

1. Kazancli E, Prchkovska V, Rodrigues P, Villoslada P, Igual L. Multiple Sclerosis Lesion Segmentation using Improved Convolutional Neural Networks. *InVISIGRAPP* (4: VISAPP) 2018 (pp. 260-269).

2. Traboulsee A, Simon JH, Stone L, et al. Revised Recommendations of the Consortium of MS Centers Task Force for a Standardized MRI Protocol and Clinical Guidelines for the Diagnosis and Follow-Up of Multiple Sclerosis. AJNR American journal of neuroradiology. 2016;37(3):394-401. doi:10.3174/ajnr.a4539

3. Kobelt G, Berg J, Atherly D, Hadjimichael O. Costs and quality of life in multiple sclerosis. Neurology. 2006;66(11):1696-1702. doi:10.1212/01.wnl.0000218309.01322.5c

4. Yang G, Yu S, Dong H, et al. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *Ieee T Med Imaging*. 2017;37(6):1310-1321. doi:10.1109/tmi.2017.2785879

5. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magnet Reson Med*. 2018;79(6):3055-3071. doi:10.1002/mrm.26977

6. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *Ieee T Med Imaging*. 2019;38(1):280-290. doi:10.1109/tmi.2018.2863670

7. Finck T, Li H, Grundl L, et al. Deep-Learning Generated Synthetic Double Inversion Recovery Images Improve Multiple Sclerosis Lesion Detection. *Invest Radiol*. 2020;55(5):318-323. doi:10.1097/rli.0000000000000640

8. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik.* 2019 May 1;29(2):102-27.

9. Zeng C, Gu L, Liu Z, Zhao S. Review of Deep Learning Approaches for the Segmentation of Multiple Sclerosis Lesions on Brain MRI. *Frontiers in Neuroinformatics.* 2020 Nov 20;14:55.

10. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. 2017;36:61-78. doi:10.1016/j.media.2016.10.004

11. Haris M, Shakhnarovich G, Ukita N. Deep Back-Projection Networks for Single Image Super-resolution. *Ieee T Pattern Anal*. 2020;PP(99):1-1. doi:10.1109/tpami.2020.3002836

12. Radue EW, Barkhof F, Kappos L, Sprenger T, Häring DA, de Vera A, von Rosenstiel P, Bright JR, Francis G, Cohen JA. Correlation between brain volume loss and clinical and MRI outcomes in multiple sclerosis. *Neurology.* 2015 Feb 24;84(8):784-93.

13. Jenkinson M, Bannister P, Brady M, Smith S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *Neuroimage*. 2002;17(2):825-841. doi:10.1006/nimg.2002.1132

14. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*. 2010;29(6):1310-1320. doi:10.1109/tmi.2010.2046908

15. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *Arxiv*. Published online 2014.

16. Hohol MJ, Orav EJ, Weiner HL. Disease Steps in multiple sclerosis: A simple approach to evaluate disease progression. Neurology 1995; 45: 251–55.

17. Hohol MJ, Orav EJ, Weiner HL. Disease Steps in multiple sclerosis: a longitudinal study comparing disease steps and EDSS to evaluate disease progression. Multiple Sclerosis 1999; 5: 349–54.

18. Marrie RA and Goldman M. Validity of performance scales for disability assessment in multiple sclerosis. Multiple Sclerosis 2007; 13: 1176-1182.

19. Wicks P, Vaughan TE, Massagli MP. The multiple sclerosis rating scale, revised (MSRS-R): development, refinement, and psychometric validation using an online community. *Health and quality of life outcomes*. 2012;10(1):70. doi:10.1186/1477-7525-10-70

20. Bove R, Secor E, Healy BC, et al. Evaluation of an online platform for multiple sclerosis research: patient description, validation of severity scale, and exploration of BMI effects on disease course. *PloS one*. 2013;8(3):e59707-e59707.

21. Wicks, P., Vaughan, T.E. & Massagli, M.P. The multiple sclerosis rating scale, revised (MSRS-R): Development, refinement, and psychometric validation using an online community. *Health Qual Life Outcomes* 10, 70 (2012). https://doi.org/10.1186/1477-7525-10-70

22. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2020. Available from: https://www.R-project.org/

23. Constance A. Mara & Robert A. Cribbie (2012) Paired-Samples Tests of Equivalence, Communications in Statistics - Simulation and Computation, 41:10, 1928-1943, DOI: 10.1080/03610918.2011.626545

24. Schuirmann, D.J. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of*

Pharmacokinetics and Biopharmaceutics 15, 657–680 (1987). https://doi.org/10.1007/BF01068419

25. Lakens D. Equivalence Tests: A Practical Primer for *t* Tests, Correlations, and Meta-Analyses. *Soc Psychol Personal Sci*. 2017;8(4):355-362. doi:10.1177/1948550617697177

26. Andrew Robinson. equivalence: Provides Tests and Graphics for Assessing Tests of Equivalence. R package version 0.7.2.; 2016. https://CRAN.R-project.org/package=equivalence

27. Wiens, B. L. (2002). Choosing an equivalence limit for noninferiority or equivalence studies. *Controlled Clinical Trials*, *23*(1), 2-14. doi:10.1016/s0197-2456(01)00196-9

28. Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. signdoi:10.18637/jss.v067.i01.

29. Mascha, Edward J. PhD*,†; Sessler, Daniel I. MD† Equivalence and Noninferiority Testing in Regression Models and Repeated-Measures Des, Anesthesia & Analgesia: March 2011 - Volume 112 - Issue 3 - p 678-687 doi: 10.1213/ANE.0b013e318206f872

30. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. *Magnet Reson Med*. 1999;42(5):952-962. doi:10.1002/(sici)1522-2594(199911)42:5<952::aid-mrm16>3.0.co;2-s

31. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnet Reson Med*. 2002;47(6):1202-1210. doi:10.1002/mrm.10171

32. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnet Reson Med*. 2007;58(6):1182-1195. doi:10.1002/mrm.21391

33. Murphy M, Alley M, Demmel J, Keutzer K, Vasanawala S, Lustig M. Fast ℓ_1-SPIRiT Compressed Sensing Parallel Imaging MRI: Scalable Parallel Implementation and Clinically Feasible Runtime. *Ieee T Med Imaging*. 2012;31(6):1250-1262. doi:10.1109/tmi.2012.2188039

34. Cheriyan J, Kim S, Wolansky LJ, Cook SD, Cadavid D. Impact of Inflammation on Brain Volume in Multiple Sclerosis. *Arch Neurol.* 2012;69(1):82–88. doi:10.1001/archneurol.2011.674

35. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. *J Gen Intern Med*. 2011;26(2):192-196. doi:10.1007/s11606-010-1513-8

36. Kleesiek J, Morshuis JN, Isensee F, et al. Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium?: A Feasibility Study. *Invest Radiol*. 2019;54(10):653-660. doi:10.1097/rli.0000000000000583

37. Narayana PA, Coronado I, Sujit SJ, Wolinsky JS, Lublin FD, Gabr RE. Deep Learning for Predicting Enhancing Lesions in Multiple Sclerosis from Noncontrast MRI. *Radiology*. 2020;294(2):398-404. doi:10.1148/radiol.2019191061