

# Influence of equipment changes on MRI measures of brain atrophy and brain microstructure in a placebo-controlled trial of ibudilast in progressive multiple sclerosis

Ken Sakaie , Janel K Fedler, Jon W Yankey, Kunio Nakamura , Josef Debbins, Mark J Lowe, Paola Raska and Robert J Fox 

Multiple Sclerosis Journal—  
Experimental, Translational  
and Clinical

April–June 2021, 1–10

DOI: 10.1177/  
20552173211010843

© The Author(s), 2021.  
Article reuse guidelines:  
[sagepub.com/journals-  
permissions](https://sagepub.com/journals-permissions)

## Abstract

**Background:** Hardware changes can be an unavoidable confound in imaging trials. Understanding the impact of such changes may play an important role in the analysis of imaging data.

**Objective:** To characterize the effect of equipment changes in a longitudinal, multi-site multiple sclerosis trial.

**Methods:** Using data from a clinical trial in progressive multiple sclerosis, we explored how major changes in imaging hardware affected data. We analyzed the extent to which these changes affected imaging biomarkers and the estimated treatment effects by including such changes as a time-dependent covariate.

**Results:** Significant differences whole brain atrophy (brain parenchymal fraction, BPF) and microstructure (transverse diffusivity, TD) between scans with and without changes were found and depended on the type of hardware change. A switch from GE HDxt to Siemens Skyra led to significant shifts in BPF ( $p < 0.04$ ) and TD ( $p < 0.0001$ ). However, we could not detect the influence of hardware changes on overall trial outcomes—differences between placebo and treatment arms in change over time of BPF and TD ( $p > 0.5$ ).

**Conclusions:** The results suggest that differences among hardware types should be considered when planning and analyzing brain atrophy and diffusivity in a longitudinal clinical trial.

**Keywords:** Multiple sclerosis, clinical trial, brain parenchymal fraction, diffusion tensor imaging, diffusivity, biomarkers

Date received: 14 December 2020; accepted: 28 March 2021

## Introduction

There is an urgent need for better imaging biomarkers for progressive multiple sclerosis (MS).<sup>1</sup> The Secondary and Primary Progressive Ibudilast NeuroNEXT Trial in Multiple Sclerosis (SPRINT-MS) was a phase II trial of ibudilast that used changes in Brain Parenchymal Fraction (BPF), a measure of whole-brain atrophy, as a primary outcome and found a significant treatment effect.<sup>2</sup> Secondary outcomes included four advanced imaging biomarkers, including diffusion tensor imaging

(DTI).<sup>3</sup> Of particular interest is transverse diffusivity (TD, also known as radial or perpendicular diffusivity). The treatment effect on TD (difference in change over time between placebo and treatment groups) was nearly 200%, larger than that of the other secondary outcomes, but was not statistically significant due to high variability. Interest in TD was driven by early work showing the correlation between TD and demyelination.<sup>4</sup> However, the correspondence between TD and diffusion

Correspondence to:

**Ken Sakaie**,  
Imaging Institute, Cleveland  
Clinic Main Campus,  
Mailcode U-15, Cleveland,  
OH 44195, USA.  
[sakaiek@ccf.org](mailto:sakaiek@ccf.org)

**Ken Sakaie**,  
Imaging Institute, The  
Cleveland Clinic, Cleveland,  
OH, USA

**Janel K Fedler**,  
Data Coordinating Center,



NeuroNEXT, University of Iowa, Iowa City, IA, USA

**Jon W Yankey,**  
Data Coordinating Center, NeuroNEXT, University of Iowa, Iowa City, IA, USA

**Kunio Nakamura,**  
Biomedical Engineering, Lerner Research Institute, The Cleveland Clinic, Cleveland, OH, USA

**Josef Debbins,**  
Keller Center for Imaging Innovation, Barrow Neurological Institute, Phoenix, AZ, USA

**Mark J Lowe,**  
Imaging Institute, The Cleveland Clinic, Cleveland, OH, USA

**Paola Raska,**  
Quantitative Health Sciences, Lerner Research Institute, Cleveland, OH, USA

**Robert J Fox,**  
Mellen Center for Multiple Sclerosis, Neurological Institute, The Cleveland Clinic, Cleveland, OH, USA

perpendicular to fiber tracts is not strict due to the presence of complex tissue geometries.<sup>5,6</sup>

The objective of this work is to assess the impact of equipment changes, a common confound in imaging-based trials. We examine the degree to which hardware changes influenced BPF, TD and the estimated effect of ibudilast. To take advantage of innovations in imaging technology, clinical radiology units routinely update scanner technology. Such updates are undesirable in trials because they may introduce systematic errors that can affect trial outcomes. We examine the extent to which such systematic errors affected the study. To contend with scanner changes, there are several common approaches. Some analyses drop scans affected by a scanner change but risk losing statistical power due to lost data. Another strategy is to consider hardware changes as a covariate in the statistical plan,<sup>7</sup> which makes it possible to indicate if details of the differences among imaging measures and among hardware need to be considered.

**Methods**

*Study overview*

SPRINT-MS was a double-blind placebo-controlled trial of ibudilast in secondary and primary progressive MS. The trial adhered to ethical guidelines in the Declaration of Helsinki.<sup>8</sup> All patients provided written informed consent. The trial included 27 scanning sites and acquired images at baseline, 24, 48, 72 and 96 weeks.<sup>9</sup> Randomization was stratified by use of immunomodulating therapy and MS diagnosis. As in the original study, analysis was confined to the 244 patients (of 255 randomized) who received at least one dose of study medication and completed one or more MRIs after baseline. Major changes in scanner hardware affected 39 subjects. Each affected subject experienced only one hardware change. The most common change (6 sites) was an upgrade from Siemens Trio to Siemens Prisma. A switch from GE HDxt to Siemens Skyra occurred at one site. One subject moved from a site with a GE HDxt to another with a GE MR750. Table 1 summarizes the prevalence of each change. Within each scanner type, major hardware components, such as head coil, are the same (Siemens Trio: a 12-channel coil. Siemens Prisma, Skyra: 20-channel coil. All GE: 8-channel coil).

*Imaging*

The nature of the imaging acquisition and postprocessing used in deriving BPF and TD differ

**Table 1.** Types and prevalence of hardware changes in the SPRINT-MS trial in terms of number of subjects affected and number of scans performed after a hardware change for treatment (Trt) and placebo (Pbo) groups, with selected demographic information and disease status.

Characteristics	GE HDxt to GE MR750		GE HDxt to Siemens Skyra		Siemens Trio to Siemens Prisma		Total Affected by Change		Total Overall	
	Trt (N = 1)	Pbo (N = 0)	Trt (N = 6)	Pbo (N = 8)	Trt (N = 12)	Pbo (N = 12)	Trt (N = 19)	Pbo (N = 20)	Trt (N = 121)	Pbo (N = 123)
No. scans	3 (2)	0	9	18 (17)	23	27 (26)	35 (34)	45 (43)	564 (560)	583 (568)
Baseline age	46.67 (n/a)	n/a	60.10 (3.90)	58.86 (3.65)	53.74 (8.02)	55.27 (7.03)	55.37 (7.55)	56.71 (6.06)	54.73 (7.71)	56.89 (6.50)
Mean (SD)										
Female sex	0	0	4	5	4	4	8	9	62	69
Primary progressive	1	0	3	5	4	7	8	12	63	64

When different from the number of scans used for BPF measurements, the number of DTI scans performed after a hardware change is in parentheses. The difference results from failure of some scans to meet quality control criteria.

substantially. Thus, the impact of hardware changes on each parameter was considered separately.

BPF is a measure of brain atrophy and has been used in many MS clinical trials and research studies.<sup>10</sup> BPF was calculated by combining T2-weighted 2D turbo/fast spin-echo images (matrix 256x256, TR/TE = 4500-6300/66-84 msec, slice thickness 3 mm, in plane resolution 1x1mm) and 2D T2-weighted FLAIR (matrix 192x256 or 256x256, TR/TE/TI = 9000-9400/77-98/2500 msec, FA = 90-120, slice thickness 3 mm, in plane resolution 1x1mm) as described in Fisher et al.,<sup>11</sup> then taking the ratio between brain parenchymal volume and outer contour volume, a smooth contour around the brain.<sup>11</sup> TD is calculated based on diffusion tensor imaging (DTI).<sup>12</sup> DTI scans were harmonized across all platforms in terms of spatial resolution and diffusion-weighting scheme (2.5 mm isotropic voxels, 255x255 x 150mm field of view, 6/8 partial Fourier factor, 64 diffusion-weighted volumes with  $b = 700 \text{ sec/mm}^2$ , 8  $b = 0$  volumes), and led to high concordance across different platforms.<sup>13</sup> All calculations were performed using software that was developed in house.<sup>14,15</sup> Further details were described previously.<sup>9,16,17</sup> and in the Appendix (supplementary material) for MRI methodology.

### Statistical analysis

The original statistical analysis plan for the SPRINT-MS trial did not adjust for hardware changes. In this analysis, we modify the original models to examine the effect of hardware changes.

A linear mixed effects model was used to estimate the rate of change in BPF by treatment group adjusted for randomization strata:

$$Y_{ij} = \alpha + \beta_{IM}(IMTherapy)_i + \beta_{MSDx}(MSDagnosis)_i + \beta_{Trt}(Treatment)_i * (time)_{ij} + \beta_{Pbo}(Placebo)_i * (time)_{ij} + a_i + b_i * (time)_{ij} + \epsilon_{ij} \quad (1)$$

where  $Y_{ij}$  is the imaging measure (BPF) for the  $i$ th subject ( $i=1, \dots, 244$ ) at the  $j$ th timepoint ( $j=0, 1, 2, 3, 4$ ),  $time_{ij}$  is the time in months, and  $\alpha$  is the common intercept.  $IMTherapy_i = 1$  (0) if the subject was (not) receiving immunomodulating therapy at randomization (interferon- $\beta$  or glatiramer acetate).  $MSDagnosis_i = 1$  (0) if the subject was diagnosed with primary (secondary) progressive MS at

randomization.  $Treatment_i = 1$  if the subject is in the treatment group, and  $Placebo_i = 1$  if the subject is in the placebo group.  $a_i$  and  $b_i$  are random effects (intercept and slope, respectively) and  $\epsilon_{ij}$  is a random error term. The model did not include a main effect of treatment and thus constrained the baseline means of the treatment groups to be equal. The test of interest compared the difference in rate of change between the treatment and placebo groups. The treatment effect is defined as the estimated difference in slopes between treatment and placebo groups ( $H_0 : \beta_{Trt} - \beta_{Pbo} = 0$  vs  $H_A : \beta_{Trt} - \beta_{Pbo} \neq 0$ ). TD was measured in two regions in each subject at each visit (left and right corticospinal tract). The model for TD was similar to that of BPF except that the model induced correlated errors between measures taken on both sides within a visit for the same subject and region was included as a fixed effect,  $\beta_{Region}(Region)_i$ , where  $Region_i = 0$  (1) for left (right) corticospinal tract:

$$Y_{ijk} = \alpha + \beta_{IM}(IMTherapy)_i + \beta_{MSDx}(MSDagnosis)_i + \beta_{Trt}(Treatment)_i * (time)_{ijk} + \beta_{Pbo}(Placebo)_i * (time)_{ijk} + \beta_{Region}(Region)_{ijk} + a_i + b_i * (time)_{ijk} + \epsilon_{ijk} \quad (2)$$

where  $Y_{ijk}$  is the imaging measure (TD) for the  $i$ th subject ( $i=1, \dots, 244$ ) at the  $j$ th timepoint ( $j=0, 1, 2, 3, 4$ ) of the  $k$ th region ( $k=1,2$ ) and  $Region_{ijk} = 0$  (1) for left (right) corticospinal tract.

The analyses of BPF and TD were modified to assess the impact of hardware changes in four ways. First, data acquired after a major hardware change were excluded from the models and the treatment effect estimated. This approach provides a simple indicator as to whether a hardware change impacted results by excluding 80 (77) data points from the BPF (DTI) model (Table 1). Second, the original model was modified to include hardware change as a time-dependent binary yes/no covariate. Equation (1) for BPF was modified to include a binary time-varying indicator,  $\beta_{ScannerChg}(ScannerChange)_{ij}$ ,

$$Y_{ij} = \alpha + \beta_{IM}(IMTherapy)_i + \beta_{MSDx}(MSDagnosis)_i + \beta_{Trt}(Treatment)_i * (time)_{ij} + \beta_{Pbo}(Placebo)_i * (time)_{ij} + \beta_{ScannerChg}(ScannerChange)_{ij} + a_i + b_i * (time)_{ij} + \epsilon_{ij} \quad (3)$$

where  $ScannerChange_{ij} = 1$  (0) if the hardware at timepoint  $j$  differed from (was the same as) that used at the baseline visit. We then estimated the overall treatment effect after adjustment. Third, the original model was altered to include type of hardware change as time-varying covariates. For BPF, the model adjusting for type of hardware change was:

$$\begin{aligned}
 Y_{ij} = & \alpha + \beta_{IM}(IMTherapy)_i \\
 & + \beta_{MSDx}(MSDiagnosis)_i \\
 & + \beta_{Trt}(Treatment)_{i*(time)_{ij}} \\
 & + \beta_{Pbo}(Placebo)_{i*(time)_{ij}} \\
 & + \beta_{GEMR750}(ScannerChangeGEMR750)_{ij} \\
 & + \beta_{SS}(ScannerChangeSiemensSkyra)_{ij} \\
 & + \beta_{SP}(ScannerChangeSiemensPrisma)_{ij} + a_i \\
 & + b_i*(time)_{ij} + \varepsilon_{ij}
 \end{aligned}
 \tag{4}$$

where  $ScannerChangeGEMR750_{ij} = 1$  if the hardware at baseline was GE HDxt and was GE MR750 at timepoint  $j$ .  $ScannerChangeSiemensSkyra_{ij} = 1$  if the hardware at baseline was GE HDxt and Siemens Skyra at timepoint  $j$ .  $ScannerChangeSiemensPrisma_{ij} = 1$  if the hardware was Siemens Trio at baseline and Siemens Prisma at timepoint  $j$ . No scanner change corresponds to  $ScannerChangeGEMR750_{ij} = 0$ ,  $ScannerChangeSiemensSkyra_{ij} = 0$  and  $ScannerChangeSiemensPrisma_{ij} = 0$ . All other covariates are defined as in equation (1). The estimated treatment effect along with the difference in imaging measure between each change type and no change was reported. Fourth, to determine whether the overall outcome varied by type of hardware change, the original model was modified to include a three-way interaction of time by treatment by type of change. Specifically, the model for BPF was:

$$\begin{aligned}
 Y_{ij} = & \alpha + \beta_{IM}(IMTherapy)_i + \beta_{MSDx}(MSDiagnosis)_i \\
 & + \beta_{TrtGEMR750}(Treatment)_{i*(time)_{ij}}*(ScannerChangeGEMR750)_{ij} \\
 & + \beta_{TrtSS}(Treatment)_{i*(time)_{ij}}*(ScannerChangeSiemensSkyra)_{ij} \\
 & + \beta_{PboSS}(Placebo)_{i*(time)_{ij}}*(ScannerChangeSiemensSkyra)_{ij} \\
 & + \beta_{TrtSP}(Treatment)_{i*(time)_{ij}}*(ScannerChangeSiemensPrisma)_{ij} \\
 & + \beta_{PboSP}(Placebo)_{i*(time)_{ij}}*(ScannerChangeSiemensPrisma)_{ij} \\
 & + \beta_{TrtNoChg}(Treatment)_{i*(time)_{ij}}*(NoScannerChange)_{ij} \\
 & + \beta_{PboNoChg}(Placebo)_{i*(time)_{ij}}*(NoScannerChange)_{ij} + a_i + b_i*(time)_{ij} + \varepsilon_{ij}
 \end{aligned}
 \tag{5}$$

where  $NoScannerChange_{ij} = 1$  if the hardware was the same at baseline and at the  $j$ th timepoint and all other covariates are defined as in equations (1) and (3). No placebo subjects had a hardware change of GE HDxt to GE MR750 thus there is no interaction term for this change and group. All model estimates are reported using restricted maximum likelihood estimation (REML).

To assess whether accounting for hardware change improved goodness-of-fit, the Akaike information criteria (AICs) were compared to the original model using Maximum Likelihood (ML) estimation. Model residuals were assessed graphically. Statistical analyses were performed in SAS 9.4 Software (SAS Institute, Cary, NC). As this is an exploratory study, no corrections for multiple comparisons were performed.

## Results

Table 2 summarizes the impact of accounting for hardware changes on overall outcomes. For BPF, accounting for hardware changes had only a slight impact on estimates of the treatment effect. For example, the original treatment effect of 0.89 (0.04, 1.74) parts per thousand per year changed to 0.84 (0.02, 1.67) parts per thousand per year after accounting for specific type of hardware change, a difference of 5.6%. For TD, accounting for hardware changes had a larger impact on the treatment effect. For example, the treatment effect went from  $-2.92$  ( $-6.85, 1.02$ )  $\times 10^{-6}$  mm<sup>2</sup>/sec per year with no adjustment to  $-3.38$  ( $-7.16, 0.41$ )  $\times 10^{-6}$  mm<sup>2</sup>/sec per year after accounting for the specific type of hardware change, a difference of 16%. For BPF and TD, adjusting for the existence of or type of hardware change led to a slight increase in precision, reflected by narrowing of the confidence intervals

**Table 2.** Effect of adjusting for hardware changes.

Measure	Model	Treatment (CI)	Placebo (CI)	Difference (CI)	p-value	AIC
BPF	Original	-0.97 (-1.57, -0.36)	-1.86 (-2.45, -1.26)	0.89 (0.04, 1.74)	0.0398	-8264.5
BPF	Exclude	-0.60 (-1.22, 0.01)	-1.53 (-2.14, -0.92)	0.93 (0.07, 1.79)	0.0348	
BPF	Binary	-0.64 (-1.25, -0.03)	-1.50 (-2.10, -0.90)	0.86 (0.02, 1.70)	0.0441	-8294.6
BPF	Type	-0.65 (-1.25, -0.05)	-1.49 (-2.08, -0.91)	0.84 (0.02, 1.67)	0.0444	-8298.6
TD	Original	-1.45 (-4.29, 1.40)	1.47 (-1.34, 4.28)	-2.92 (-6.85, 1.02)	0.1465	-10006.5
TD	Exclude	-1.23 (-4.14, 1.68)	2.51 (-0.40, 5.42)	-3.74 (-7.81, 0.32)	0.0711	
TD	Binary	-0.68 (-3.56, 2.19)	2.38 (-0.48, 5.24)	-3.07 (-6.97, 0.84)	0.1234	-10012.4
TD	Type	-0.80 (-3.59, 1.98)	2.57 (-0.20, 5.34)	-3.38 (-7.16, 0.41)	0.0802	-10033.9

Values associated with BPF are rate of change (slope) and have units of parts per thousand per year. 95% confidence intervals are given in parentheses. Values associated with TD are rate of change (slope) and have units of  $10^{-6}$  mm<sup>2</sup>/sec per year. p-values refer to the test of the difference in rate of change between the treatment and placebo groups. In this study, a year is defined as 48 weeks. Models are the version used in Fox et al.<sup>2</sup> with no adjustment for hardware changes (Original), in which data acquired after a hardware change are excluded (Exclude), in which hardware change is treated as a binary yes/no time-dependent covariate (Binary), and in which type of hardware change is a time-dependent covariate (Type). All model estimates use REML estimation. AIC is reported based on ML estimation. A difference in AIC of 2 indicates improved model fit with lower AIC being better. AIC is not reported for the Exclude case because AIC is not comparable between models using different outcome data points.



**Table 3.** Effect of hardware change on BPF and TD.

Type	BPF (CI)	p-value	TD (CI)	p-value
GE HDxt to GE MR750	4.23 (−4.29, 12.74)	0.3299	37.22 (−5.60, 80.05)	0.0884
GE HDxt to Siemens Skyra	−2.36 (−4.54, −0.18)	0.0340	−30.99 (−41.94, −20.04)	<0.0001
Siemens Trio to Siemens Prisma	−5.27 (−7.00, −3.55)	<0.0001	0.99 (−7.63, 9.62)	0.8218

BPF is reported in parts per thousand. TD is in units of  $\times 10^{-6}$  sec/mm<sup>2</sup>. 95% confidence intervals are given in parentheses. Values are estimates of the shift caused by each type of hardware change versus no change. p-values are for the test of the estimate of the scanner change-induced shift being equal to zero.

of the treatment effect. Values of AIC suggest that adjusting for hardware changes improved the model goodness-of-fit. The model adjusting for scanner upgrade type, which allowed for each scanner upgrade type to have a different effect, was optimal.

Closer examination of the covariate analysis shows that different types of hardware changes have qualitatively different behavior. Table 3 shows the impact of type of hardware change on imaging measures. Values are systematic differences between imaging measures due to a particular type of hardware change. The change from Siemens Trio to Siemens Prisma is associated with a systematic lowering of BPF. Among those affected by this change, an equal number were in the treatment and placebo arms (12 each, Table 1). The number of scans within each arm was nearly equal—23 (27) in the treatment (placebo) arm. This balance between may explain why the systematic shift in BPF did not affect the overall outcome. In contrast, the change from GE HDxt to Siemens Skyra, associated with a systematic lowering of TD, affected nearly twice as many scans within the placebo (17) than in the treatment arm (9). Together, the systematic shift and the imbalance could explain why the treatment effect is larger after accounting for the hardware change. To understand why, consider that TD decreases over time within the treatment arm, as can be seen by the negative values of slope in Table 2, while TD increases over time within the placebo arm, as can be seen by the positive values of slope. A hardware change will affect later time points, and a systematic reduction in TD will decrease the slope of change over time (less positive slope for placebo and more negative slope for treatment). Such a trend is exemplified within the placebo arm, Table 2. The slope is smaller without adjustment for hardware changes ( $1.47$  (−1.34, 4.28)  $\times 10^{-6}$  mm<sup>2</sup>/sec per year) than with adjustment ( $2.38$  (−0.48, 5.24) to  $2.57$  (−0.20, 5.34)

$\times 10^{-6}$  mm<sup>2</sup>/sec per year). In the treatment arm, the effect of hardware changes follows the same pattern ( $−1.45$  (−4.29, 1.40)  $\times 10^{-6}$  mm<sup>2</sup>/sec per year without adjustment,  $−0.68$  (−3.56, 2.19) to  $−1.23$  (−4.14, 1.68)  $\times 10^{-6}$  mm<sup>2</sup>/sec per year with adjustment), but is less pronounced, perhaps reflecting the lower number of scans affected by the hardware change within the treatment arm.

Scanner changes affected imaging measures, but there was no indication that the effect of treatment varied by scanner upgrade type. Analysis including a time by treatment by type of change term found no significant interaction between type of scanner change and treatment effect for BPF or TD (Table 4,  $p > 0.5$ ). However, this study was probably not sufficiently powered to draw a conclusion from this result.

Figure 1 illustrates the findings for BPF and a scanner change from a Siemens Trio to a Siemens Prisma at week 72. The scanner change led to a negative shift in BPF values. The magnitude of the rate of change was overestimated when hardware changes were ignored as in the original models. However, accounting for hardware changes had little impact on the treatment effect, the difference in slopes between treatment groups.

Figure 2 focuses on TD and the effect of a scanner change of GE HDxt to Siemens Skyra. Here the estimated change over time within the treatment (placebo) groups was underestimated in the original model not adjusting for scanner changes. The magnitude of the shift due to scanner change is large, but had only a slight impact on the overall treatment effect.

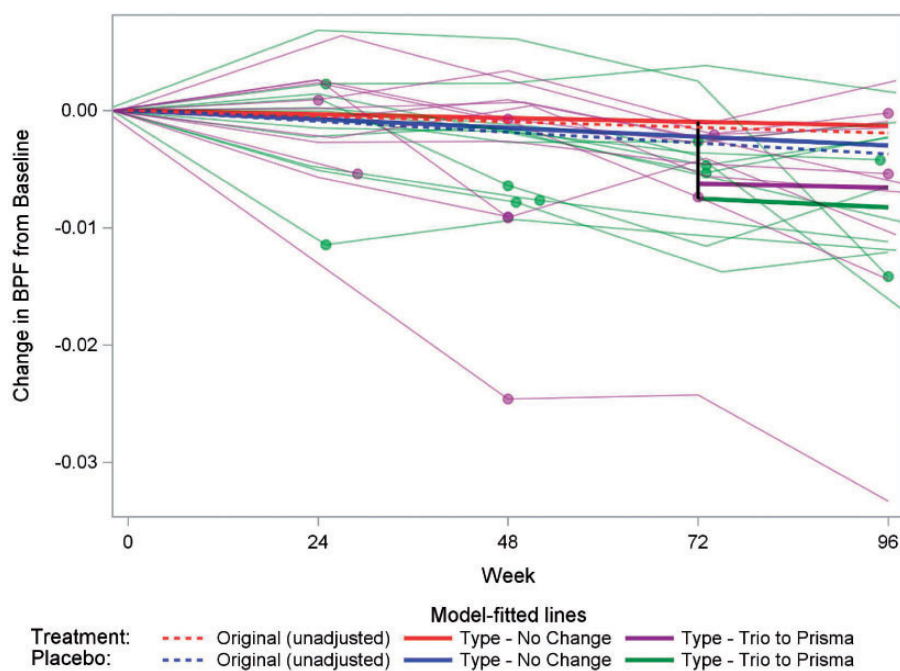
### Discussion

This analysis found that hardware changes affected estimates of atrophy and diffusivity but had little

**Table 4.** Impact of scanner change on overall outcomes.

Measure	Type	Difference treatment effect change vs no change	p-value
BPF	GE HDxt to Siemens Skyra	0.52 (−1.94, 2.99)	0.6758
BPF	Siemens Trio to Siemens Prisma	0.17 (−1.88, 2.23)	0.8698
TD	GE HDxt to Siemens Skyra	−1.66 (−14.05, 10.72)	0.7924
TD	Siemens Trio to Siemens Prisma	3.44 (−0.61, 1.23)	0.5056

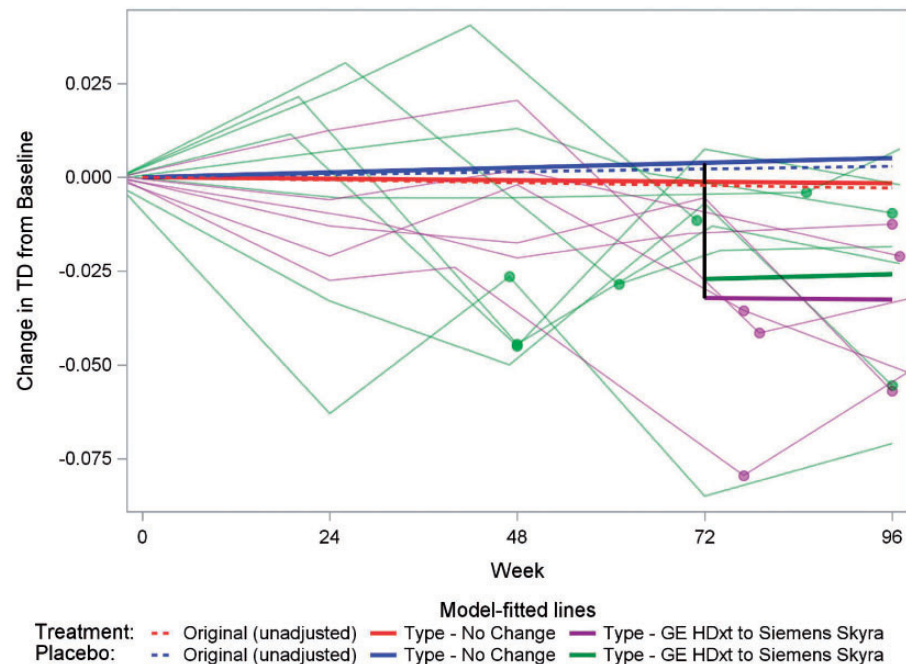
Values are differences in treatment effect for BPF (parts per thousand) or TD ( $\times 10^{-6}$  sec/mm<sup>2</sup>) over 48 weeks by a particular type of hardware change versus no change. p-values are for the test of this difference. The GE HDxt to GE MR750 change could not be analyzed because only one subject was affected.



**Figure 1.** Illustration of results for BPF and the effect of one of the three common types of scanner change (Siemens Trio to Siemens Prisma). The dashed red (blue) line is the result from the model for the treatment (placebo) group in the original model, in which there was no adjustment for scanner change. The solid red (blue) line is the estimate for the treatment (placebo) group assuming no scanner change from the model that adjusts for type of scanner change. The values of the slopes of the red and blue lines are given in Table 2. The vertical black line indicates the shift in values associated with an example scanner change occurring at week 72 (Table 3). The solid thick purple (green) line is the estimated change over time for the treatment (placebo) group after the example scanner change at week 72 from the model that adjusts for type of scanner change. The solid red (blue) and purple (green) lines are parallel; the effect of the upgrade is illustrated by the shift down (black line). The solid thin purple (green) lines show individual patients' data for the subset of treatment (placebo) patients who experienced a scanner change from a Siemens Trio to a Siemens Prisma. The dots indicate the time of the first scan after the scanner change for a patient (not necessarily occurring at week 72). Sensitivity analyses excluding the outlying subject (purple line at bottom) were considered and had no effect on the conclusions drawn.

impact on the overall outcomes of the trial. The direction and magnitude differed among type of hardware change and between imaging-based measures. Retrospective analyses of the effect of scanner changes have been performed,<sup>7,18</sup> but typically focus

on a single imaging measure. Our results highlight the need to consider different imaging measures separately and to account for the type of hardware change. Although BPF and TD are measured on the same scanner at each visit, the nature of the



**Figure 2.** Illustration of results for TD and the effect of a scanner change of one of the three common types of scanner change (GE HDxt to Siemens Skyra). The dashed red (blue) line is the result from the model for the treatment (placebo) group in the original model, in which there was no adjustment for scanner change. The solid red (blue) line is the estimate for the treatment (placebo) group assuming no scanner change from the model that adjusts for type of scanner change. The values of the slopes of the red and blues lines are given in Table 2. The vertical black line indicates the shift in values associated with an example scanner change occurring at week 72 (Table 3). The solid thick purple (green) line is the estimated change over time for the treatment (placebo) group after the example scanner change at week 72 from the model that adjusts for type of scanner change. The solid red (blue) and purple (green) lines are parallel; the effect of the upgrade is illustrated by the shift down (black line). The solid thin purple (green) lines show individual patients' data for the subset of treatment (placebo) patients who experienced a scanner change from a GE HDxt to Siemens Skyra. Individual patient TD values were averaged over the left and right side when plotting. The dots indicate the time of the first scan after the scanner change for a patient (not necessarily occurring at week 72).

measures differs and thus are differently affected by scanner changes. Different behavior may be seen among other secondary measures, which will be examined in future work.

The study design can offset some of the effect of equipment changes. If equipment changes are randomly distributed between treatment and placebo arms, occur relatively infrequently and largely affect shifts in the outcome, the consequences may be minimal (Table 2) even if the shift and associated variability due to hardware change is large (Table 3). The shifts associated with hardware changes (Table 3) are an order of magnitude larger than the annual changes (Table 2), but accounting for these changes had only small impacts on treatment effect. Assessments of longitudinal change within a treatment group and single-arm trials may be more susceptible to biases introduced by equipment changes. A sensitivity analysis could be performed to determine the degree to which

imbalances in equipment changes might affect the results.

Much work has investigated imaging hardware associated differences. Prospective measurement can determine if systematic differences are smaller than physiological differences of interest.<sup>19,20</sup> Phantom measurements can be used to measure differences<sup>17,20–23</sup> and can help minimize systematic differences prospectively,<sup>24</sup> but may not relate to *in vivo* measurements.<sup>25</sup> It may be difficult to access newly-released hardware for prospective measurements. Retrospective harmonization<sup>26,27</sup> is an active field but has not, to our knowledge, been extensively tested in a randomized controlled trial.

A number of other factors have been explored. Retrospective analysis of the SPRINT-MS data showed that the time by treatment effect differs between the secondary and primary progressive groups.<sup>28</sup> We used a four-way interaction of time,



treatment group, MS diagnosis and scanner upgrade to explore if the treatment effect differed by scanner upgrade within each MS diagnosis group. We did not detect an effect, but this analysis is likely to be underpowered. We reported differences arising from hardware types among healthy controls<sup>16</sup> and variances may differ among hardware types but such differences were not included in the original models. In an exploratory analysis, we adjusted for baseline differences in acquisition systems by adding a categorical variable for baseline scanner type in equation (4). We also investigated differences in variance in the imaging outcomes across scanner model by specifying scanner model-specific residual variance.<sup>29</sup> This was done by modifying the covariance parameters of equation (4) to allow heterogeneous residual variance across scanner models in addition to the random intercept and slope. Neither of these analyses resulted in substantial changes in the results.

### Conclusions

Because imaging hardware changes can be expected in imaging-based clinical trials, anticipating the impact of such changes on the trial outcomes and adjusting the analysis plan is prudent. Accounting for the hardware change can be important, but the specific type of imaging metric and hardware change should be considered.

### Declaration of Conflicting Interests


The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: KS has received salary support from Genzyme and Novartis. RJF has received personal consulting fees from AB Science, Actelion, Biogen, Celgene, EMD Serono, Genentech, Immunic, Novartis, Sanofi, Teva, and TG Therapeutics, has served on advisory committees for Actelion, Biogen, Immunic, Novartis, and Sanofi, and has received clinical trial contract and research grant funding from Biogen and Novartis. JKF, JWY, KN, JD, MJL and PR have no relevant conflicts to disclose.


### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the National Institute of Neurological Disorders and Stroke (U01NS082329) and the National Multiple Sclerosis Society (RG 4778-A-6) and by MediciNova through a contract with the National Institutes of Health.

### ORCID iDs

Ken Sakaie  <https://orcid.org/0000-0002-5633-4494>

Kunio Nakamura  <https://orcid.org/0000-0002-7833-8138>

Robert J Fox  <https://orcid.org/0000-0002-4263-3717>

### Supplemental Material

Supplemental material for this article is available online.

### References

- Oh J, Ontaneda D, Azevedo C, et al.; North American Imaging in Multiple Sclerosis Cooperative. Imaging outcome measures of neuroprotection and repair in MS: a consensus statement from NAIMS. *Neurology* 2019; 92: 519–533.
- Fox RJ, Coffey CS, Conwit R, et al.; NN102/SPRINT-MS Trial Investigators. Phase 2 trial of ibudilast in progressive multiple sclerosis. *N Engl J Med* 2018; 379: 846–855.
- Basser PJ, Pajevic S, Pierpaoli C, et al. In vivo fiber tractography using DT-MRI data. *Magn Reson Med* 2000; 44: 625–632.
- Budde MD, Kim JH, Liang HF, et al. Toward accurate diagnosis of white matter pathology using diffusion tensor imaging. *Magn Reson Med* 2007; 57: 688–695.
- Jeurissen B, Leemans A, Tournier JD, et al. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Hum Brain Mapp* 2013; 34: 2747–2766.
- Wheeler-Kingshott CA and Cercignani M. About “axial” and “radial” diffusivities. *Magn Reson Med* 2009; 61: 1255–1260.
- Lee H, Nakamura K, Narayanan S, et al.; Alzheimer’s Disease Neuroimaging Initiative. Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements. *Neuroimage* 2019; 184: 555–565.
- World Medical Association. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013; 310: 2191–2194.
- Fox RJ, Coffey CS, Cudkovic ME, et al. Design, rationale, and baseline characteristics of the randomized double-blind phase II clinical trial of ibudilast in progressive multiple sclerosis. *Contemp Clin Trials* 2016; 50: 166–177.
- Rudick RA, Fisher E, Lee JC, et al.; Multiple Sclerosis Collaborative Research Group. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. *Neurology* 1999; 53: 1698–1704.
- Fisher E, Cothren R, Tkach J, et al. Knowledge-based 3D segmentation of the brain in MR images for quantitative multiple sclerosis lesion tracking. *SPIE Med Imag* 1997; 3034, 19–25.
- Basser PJ, Mattiello J and LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B* 1994; 103: 247–254.
- Fox RJ, Sakaie K, Lee JC, et al. A validation study of multicenter diffusion tensor imaging: reliability of fractional anisotropy and diffusivity values. *AJNR Am J Neuroradiol* 2012; 33: 695–700.

14. Nakamura K. *MRI analysis to detect gray matter tissue loss in multiple sclerosis*. Department of Biomedical Engineering, Case Western Reserve University, 2011, p. 189.
15. Lowe MJ, Beall EB, Sakaie KE, et al. Resting state sensorimotor functional connectivity in multiple sclerosis inversely correlates with transcallosal motor pathway transverse diffusivity. *Hum Brain Mapp* 2008; 29: 818–827.
16. Zhou X, Sakaie KE, Debbins JP, et al. Scan-rescan repeatability and cross-scanner comparability of DTI metrics in healthy subjects in the SPRINT-MS multicenter trial. *Magn Reson Imaging* 2018; 53: 105–111.
17. Zhou X, Sakaie KE, Debbins JP, et al. Quantitative quality assurance in a multicenter HARDI clinical trial at 3T. *Magn Reson Imaging* 2017; 35: 81–90.
18. Stonnington CM, Tan G, Kloppel S, et al. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage* 2008; 39: 1180–1185.
19. Sutton BP, Goh J, Hebrank A, et al. Investigation and validation of intersite fMRI studies using the same imaging hardware. *J Magn Reson Imaging* 2008; 28: 21–28.
20. Duchesne S, Chouinard I, Potvin O, et al.; for the CIMA-Q group and the CCNA group. The Canadian dementia imaging protocol: harmonizing national cohorts. *J Magn Reson Imaging* 2019; 49: 456–465.
21. Keenan KE, Gimbutas Z, Dienstfrey A, et al. Assessing effects of scanner upgrades for clinical studies. *J Magn Reson Imaging* 2019; 50: 1948–1954.
22. Fu L, Fonov V, Pike B, et al. *Automated analysis of multi site MRI phantom data for the NIHPD project*. Berlin: Springer, 2006, pp. 144–151.
23. Friedman L and Glover GH. Report on a multicenter fMRI quality assurance protocol. *J Magn Reson Imaging* 2006; 23: 827–839.
24. Keenan KE, Biller JR, Delfino JG, et al. Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. *J Magn Reson Imaging* 2019; 49: e26–e39.
25. Wilde EA, Bigler ED, Huff T, et al. Quantitative structural neuroimaging of mild traumatic brain injury in the chronic effects of neurotrauma consortium (CENC): comparison of volumetric data within and across scanners. *Brain Inj* 2016; 30: 1442–1451.
26. Cetin Karayumak S, Bouix S, Ning L, et al. Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters. *Neuroimage* 2019; 184: 180–200.
27. Fortin JP, Parker D, Tunc B, et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 2017; 161: 149–170.
28. Goodman AD, Fedler JK, Yankey J, et al.; the SPRINT-MS Investigators. Response to ibudilast treatment according to progressive multiple sclerosis disease phenotype. *Ann Clin Transl Neurol* 2021; 8: 111–118.
29. Chua AS, Egorova S, Anderson MC, et al. Handling changes in MRI acquisition parameters in modeling whole brain lesion volume and atrophy data in multiple sclerosis subjects: Comparison of linear mixed-effect models. *Neuroimage Clin* 2015; 8: 606–610.