University of Pennsylvania
UPenn Biostatistics Working Papers

Year 2015 *Paper* 37

Statistical Estimation of T1 Relaxation Time Using Conventional Magnetic Resonance Imaging

Amanda Mejia* Elizabeth M. Sweeney† Blake Dewey‡
Govind Nair** Pascal Sati†† Colin Shea‡‡
Daniel S. Reich§ Russell T. Shinohara¶

*Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, amejia@jhsph.edu

†Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University; Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, emsweene@jhsph.edu

‡Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health

**Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health

††Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health

‡‡Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health

§Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health; Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, reichds@ninds.nih.gov

¶Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, rshi@upenn.edu

Statistical Estimation of T1 Relaxation Time Using Conventional Magnetic Resonance Imaging

Amanda Mejia, Elizabeth M. Sweeney, Blake Dewey, Govind Nair, Pascal Sati, Colin Shea, Daniel S. Reich, and Russell T. Shinohara

Abstract

Quantitative T_1 maps estimate T_1 relaxation times and can be used to assess diffuse tissue abnormalities within normal-appearing tissue. T_1 maps are popular for studying the progression and treatment of multiple sclerosis (MS). However, their inclusion in standard imaging protocols remains limited due to the additional scanning time and expert calibration required and susceptibility to bias and noise. Here, we propose a new method of estimating T_1 maps using four conventional MR images, which are intensity- normalized using cerebellar gray matter as a reference tissue and related to T_1 using a smooth regression model. Using leave-one-out cross-validation, we generate statistical T_1 maps for 61 subjects with MS. The statistical maps are less noisy than the acquired maps and show similar accuracy. Tests of group differences in normal-appearing white matter across MS subtypes give similar results using both methods, but tests performed using statistical maps are more powerful.

Statistical estimation of T_1 relaxation times using conventional magnetic resonance imaging

Amanda Mejia[1], Elizabeth M Sweeney[1,2], Blake Dewey[2], Govind Nair[2], Pascal Sati[2], Colin Shea[2], Daniel S Reich[1,2] & Russell T Shinohara[3*]

[1] Department of Biostatistics, The Johns Hopkins University, Baltimore, MD 21205

[2] Translational Neuroradiology Unit, Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disease and Stroke, National Institute of Health, Bethesda, MD 20892

[3] Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Abstract

Quantitative T_1 maps estimate T_1 relaxation times and can be used to assess diffuse tissue abnormalities within normal-appearing tissue. T_1 maps are popular for studying the progression and treatment of multiple sclerosis (MS). However, their inclusion in standard imaging protocols remains limited due to the additional scanning time and expert calibration required and susceptibility to bias and noise. Here, we propose a new method of estimating T_1 maps using four conventional MR images, which are intensity-normalized using cerebellar gray matter as a reference tissue and related to T_1 using a smooth regression model. Using leave-one-out cross-validation, we generate statistical T_1 maps for 61 subjects with MS. The statistical maps are less noisy than the acquired maps and show similar accuracy. Tests of group differences in normal-appearing white matter across MS subtypes give similar results using both methods, but tests performed using statistical maps are more powerful.

*Correspondence should be addressed to R.T.S. (rshi@upenn.edu).

1 Introduction

Quantitative magnetic resonance imaging (MRI) techniques are used to estimate features of human body tissue for the study of progression and treatment of diseases. For example, T_1 maps estimate T_1 relaxation times, which can be used as an indicator of inflammation and in nervous system tissue, demyelination, axonal loss and gliosis [1, 2]. Increased relaxation times across normal-appearing white matter of the brain have been associated with a number of neurological diseases and disorders, including multiple sclerosis (MS), schizophrenia, alcoholism, optic neuritis, and near-terminal AIDS [3].

In MS specifically, elevated T_1 relaxation times have been consistently observed in normal-appearing white matter (NAWM), relative to that of healthy control subjects [1, 4, 5, 6, 7]. There is also evidence for elevated T_1 in MS patients in deep gray matter (DGM), particularly in the thalamus [1, 4, 6], and in cortical gray matter (CGM) [1]. Longitudinal studies have shown that the variance of T_1 in NAWM and CGM increases over time for MS patients [8, 9, 10, 11] and that average T_1 in NAWM may also increase [10]. Furthermore, T_1 is a clinically meaningful measure that is necessary to obtain a full picture of MS disease burden. In particular, the mean and variance of T_1 in NAWM have been found to be predictive of current and future disability [6, 10], and elevated T_1 in DGM has been associated with fatigue [12, 13]. Compared with lesion-based assessments of MS disease burden, which can be detected on standard clinical MR images, T_1 maps appear to provide complementary, not redudant, information [4, 11].

However, multiple factors have limited the widespread availability of T_1 maps. First, T_1 maps are still not included in most standard clinical or research protocols, due in part to the additional scanning time required for their acquisition. This is particularly true in the clinical setting, where limited resources and patient considerations typically limit acquisition to conventional MR images. Second, while the availability of T_1 maps in observational studies is increasing, many longitudinal studies of MS and other neurological diseases include years or even decades of imaging history, during only a fraction of which T_1 maps may have been acquired, thus limiting our longitudinal understanding of T_1 in these diseases. Finally, T_1 mapping requires very careful scanner calibration, and many technical and environmental factors can introduce bias and noise, limiting reproducibility across acquisition methods, centers, scanners or visits.

In this paper, we introduce a novel method of computing T_1 maps. Traditional T_1 maps estimate T_1 analytically, using multiple (often two) T_1w images acquired with different flip angles or inversion times to interpolate the T_1 relaxation curve. This curve is a non-linear function of T_1, a constant at a given physical location, and can be back-solved to estimate T_1. By contrast, our method estimates T_1 statistically using a voxel-wise regression model based on four conventional MR images that are already included in many standard protocols: T_1-weighted (T_1w), T_2-weighted (T_2w), proton density-weighted (PDw), and T_2-weighted fluid attenuated inversion recovery ($FLAIR$). Using our method, a T_1 map can be added to a clinical or research study at any point after acquiring these four images.

A novel intensity normalization technique that utilizes the cerebellar gray matter (CBGM) as a reference tissue is key to the success of our method, Quantitative MR Estimation Employing Normalization (QuEEN). The cerebellum has been previously utilized as a reference region for intensity normalization of positron emission tomography (PET) in the contexts of MS [14] and Alzheimer's disease [15], though to the best of our knowledge it has not been used for intensity normalization of MRI. We use cerebellar gray matter only, rather than the entire cerebellum, since the latter is sensitive to the relative volume of gray and white matter within each subject's cerebellum. Henceforth, we refer to the traditional, analytically estimated T_1 maps as "T_1 maps" and to the QuEEN-estimated T_1 maps as "QuEEN maps". The QuEEN model is fully automated and computationally efficient.

2 Methods

In this section, we detail the QuEEN method of T_1 estimation. QuEEN is comprised of two primary steps: intensity normalization of four predictor MR images, followed by training of a regression model

relating the intensity-normalized predictor images to T_1. Before either of these steps takes place, a standard MRI preprocessing pipeline is performed, including rigid registration (alignment) of images within each subject, removal of voxels corresponding to extracerebral tissue, skull and background, and segmentation of brain tissue types. We validate the proposed methods by demonstrating that QuEEN maps are comparable to T_1 maps in terms of accuracy and utility. Most notably, we establish that QuEEN maps can detect group differences in acute and diffuse white matter pathologies between subjects with different MS subtypes as well as, and in some cases better than, T_1 maps. Finally, we show that QuEEN and T_1 maps have similar measurement error rates.

2.1 QuEEN Model

2.1.1 Intensity Normalization

The four MR images used as predictors in the QuEEN model are unitless, and therefore their intensities only have relative meaning within a scan. Intensity normalization of conventional MR images is often performed to make intensities comparable across images by converting intensities to deviations from some reference point. The z-score normalization method [16, 17] subtracts an estimate of the "center" of the distribution of intensities in an image and divides by a measure of its "scale". An appropriate choice of reference region(s) to estimate this center and scale can uncover quantitative information within the predictor images (e.g. elevated T_1 in NAWM of MS patients) and is crucial for accurate estimation of T_1 in the QuEEN model.

Unfortunately, two commonly used reference regions—normal appearing white matter (NAWM) [16] and cerebral spinal fluid (CSF) [18, 19, 20, 21, 22, 23]—can introduce unnecessary bias or variance into the estimates of T_1 in the QuEEN model. We argue that use of CSF as a reference class would lead to a variance problem, while use of NAWM would lead to a bias problem. T_1 of CSF is generally unaffected by disease, making it a potentially attractive reference class. However, MRI intensities within CSF tend to contain high levels of noise, which introduces variance into intensity-normalized MR images. On the other hand, the distribution of intensities in NAWM is well-estimated, making it a stable reference class. However, in MS patients and other disease populations, diffuse changes in T_1 within NAWM are commonplace and furthermore are of interest to researchers. Intensity-normalizing with respect to NAWM would obscure such changes, as normalized intensities measure deviation from the reference class.

Therefore, we propose a novel intensity normalization method that utilizes a combination of cerebellar gray matter (CBGM) and NAWM. This method takes advantage of qualities of both tissue classes: T_1 in CBGM is similar across subjects and disease groups, making it a meaningful "center"; NAWM is well-estimated, making it a stable reference class for estimation of "scale".

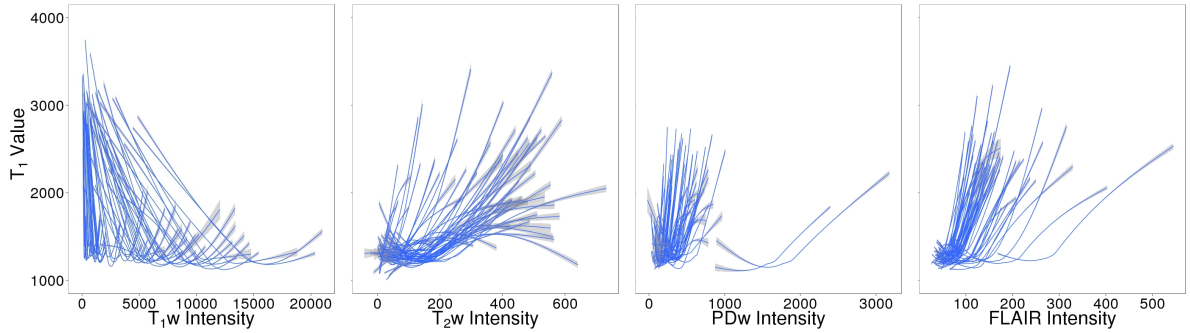
Let $M_i(v)$ denote the raw intensity of voxel v for subject i in image $M \in \{FLAIR, PDw, T_1w, T_2w\}$. Using the z-score method, we normalize image M_i:

$$M_i^N(v) = \frac{M_i(v) - \mu_{i,M}^{(CBGM)}}{\sigma_{i,M}^{(NAWM)}},$$

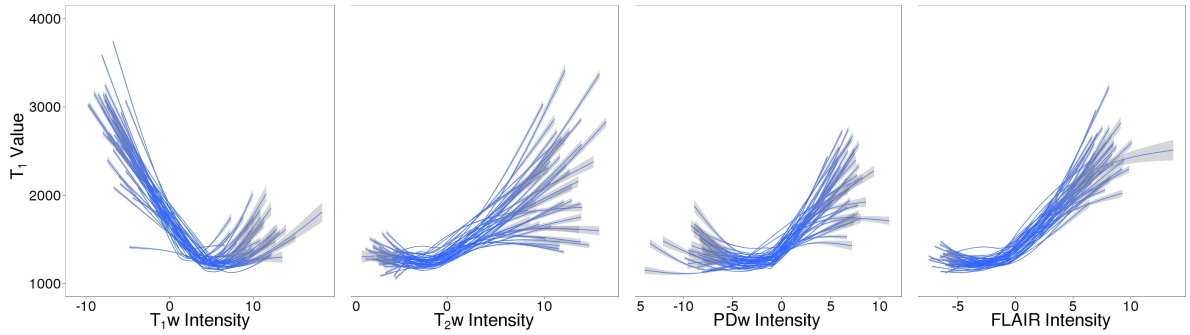
where $\mu_{i,M}^{(CBGM)}$ is the median intensity within CBGM and $\sigma_{i,M}^{(NAWM)}$ is the standard deviation of intensities within NAWM. Figure 1 shows the relationships between T_1 and each predictor image before and after normalization and illustrates how normalization serves to standardize these curves across subjects.

2.1.2 Statistical Model

Following intensity normalization, a set of nonlinear regression models is trained to predict T_1 from the normalized predictor images. A separate regression model is fit within each tissue class. We assume independence across subjects; furthermore we assume that, conditional on the predictor images, T_1 is independent across voxels. The T_1 map intensity at voxel v in tissue class c for subject i is therefore



(a) T_1 versus *non-normalized* intensities of T_1w, T_2w, PDw and $FLAIR$ within NAWM



(b) T_1 versus *normalized* intensities of T_1w, T_2w, PDw and $FLAIR$ within NAWM

Figure 1: T_1 versus *non-normalized* (a) and *normalized* (b) intensities of predictor images within NAWM. Each line is a smoothed curve from a single subject, and confidence bands for each curve are shown in gray. Confidence bands are wider for more extreme values of the normalized predictor images, due to low density of voxels with such intensities. Before intensity normalization, the relationships between T_1 and each predictor image varies greatly across subjects; after intensity normalization, the curves appear much more similar across subjects. Furthermore, the relationships between T_1 and each predictor image are clearly nonlinear.

related to the corresponding normalized T_1w, T_2w, PDw and $FLAIR$ intensities through the following model:

$$T_{1i}(v) \sim f_1^c\left(T_1w_i^N(v)\right) + f_2^c\left(T_2w_i^N(v)\right) + f_3^c\left(PDw_i^N(v)\right) + f_4^c\left(FLAIR_i^N(v)\right) + \epsilon_i(v),$$

where $f_j^c(\cdot)$, $j = 1, 2, 3, 4$, are smooth curves and $\epsilon_i(v) \sim N(0, \sigma_c^2)$. We fit this model in the R statistical environment (version 3.0.2, [24]) using the `gam` function from the `mgcv` package (version 1.7-28, [25, 26]). This function represents the smooth curves as penalized regression splines. Generalized cross validation is used to estimate the degree of smoothness, and the smoothing parameter estimation criterion is optimized using the Newton method [27]. For any subject with T_1w, T_2w, PDw and $FLAIR$ images, a QuEEN map may be obtained by applying the estimated regression curves $\hat{f}_j^c(\cdot)$, $j = 1, 2, 3, 4$, to the corresponding normalized predictor image intensities.

2.2 Method Validation

We assess the performance of the QuEEN model using several criteria. To assess the *accuracy* of QuEEN maps, we compute their *estimation error* and *prediction error*. We define the estimation error of a QuEEN

map at each voxel as the difference between the intensities on the QuEEN map and corresponding T_1 map. We note that the regression framework used to generate QuEEN maps may serve to reduce the noise typically found on T_1 maps due to the "shrinkage" effect of regression, in which observations are pulled towards the mean. The shrinkage effect has previously been shown to improve reproducibility of functional MRI measures [28, 29]. However, it is difficult to determine what constitutes a good or bad estimation error without knowing the measurement error of T_1 maps.

Therefore we also define prediction error, the difference at each voxel between a T_1 or QuEEN map and a second T_1 map acquired at a later date. We compute the prediction error of the T_1 and QuEEN maps of those subjects in our dataset for whom a second T_1 map is available. Note that while prediction error may be affected by real biological changes, such changes will affect the prediction errors of T_1 and QuEEN maps similarly, and hence they can be compared fairly. For each error rate, we compute the root mean squared error (rMSE) within each image and tissue class and compare the distributions of rMSE across subjects. It is important to note that, due to measurement error inherent in T_1 maps and the possibility for such errors to be recurring across multiple scans, both error measures of QuEEN maps may be somewhat inflated.

To assess *utility* of QuEEN maps, we assess the performance of QuEEN maps in detecting acute and diffuse changes in white matter, a common use of T_1 maps. Specifically, we perform tests of group differences of T_1 in NAWM and lesions between different MS subtypes using both T_1 and QuEEN maps and compare the results. Previous work suggests that SPMS subjects show elevated T_1 in NAWM relative to RRMS and PPMS patients, and that RRMS patients show elevated T_1 in NAWM relative to PPMS patients [1]. Therefore, for each of these pairs we conduct a one-sided test for differences in T_1 in NAWM. Much extant literature also finds that patients with MS show elevated T_1 relative to healthy volunteers (HVs) [1, 4, 5, 6, 8, 30]. We therefore also conduct a one-sided test for differences in T_1 in NAWM for MS patients and each MS subtype versus HVs. However, as our HV group is very small, we present these results as preliminary findings in Appendix Figure D.6. Finally, we also perform a two-sided test for differences in T_1 in lesions between each pair of MS subtypes.

2.3 Materials

2.3.1 Study Population

Our dataset consists of MRI studies collected from 75 subjects. In order to ensure image quality for model training and validation, we performed extensive quality control. Four studies were excluded due to subject motion. An additional 2 studies were excluded due to registration problems, and 6 studies were excluded due to tissue class segmentation errors. Of the 63 remaining studies, 29 are from patients with PPMS, 15 are from patients with RRMS, 17 are from patients with SPMS, and 2 are healthy volunteers (HVs). Additional summary statistics are shown in Table 1. For model training, we formed a high-quality dataset containing 45 studies by excluding any studies where even minor segmentation errors or subject motion were present. The high-quality dataset was also included in model validation, using cross-validation for image prediction within this dataset to avoid over-fitting.

To assess reproducibility, a second MRI study was collected for 37 subjects. Of these, one study was excluded due to motion, and two studies were excluded due to segmentation errors. Of the remaining 34 studies, 20 are from patients with PPMS, 13 are from patients with SPMS, and one is from a patient with RRMS. The average length of time between the two studies is 169 days and ranges from 21 to 301 days.

2.3.2 Image Acquisition

Each MRI study includes the following images, all collected on a Siemens Skyra 3T scanner: a T_1 map, acquired as two T_1-FLASH (Fast Low Angle SHot) sequences at differing flip angles [TR=7.8ms, TE=3ms, FA=3/18] [31, 32] with a B0+B1 field map for FA correction [33]; T_1-MPRAGE (Magnetization-Prepared

Table 1: Summary statistics of study population

	HV	PPMS	RRMS	SPMS
n	2	29	15	17
% Female	50%	55%	87%	41%
Mean Age (sd)	24 (0.6)	56 (7.4)	45 (13.8)	53 (7.9)
Mean Disease Duration (sd)	NA	13.6 (9.4)	8.2 (6.8)	24.6 (8.8)
Median EDSS (range)	NA	6.0 (2.0-7.5)	1.5 (1.0-6.0)	6.5 (1.5-7.0)

RApid Gradient Echo) [TR=3000ms, TE=3.03ms, TI=900ms, FA=9]; PDw and T_2w images from a dual-echo turbo spin echo (TSE) sequence [TR=3000ms, TE=11ms/101ms, FA=150, ETL=14]; and a 3D T_2-weighted $FLAIR$ image acquired using a T_2-selective inversion pulse optimized for T_2 of 120ms [TR=4800ms, TE=354ms, TI=1800ms, Variable FA]. All scans were acquired at 1.0mm isotropic resolution except the PDw/T_2w TSE sequence, which was acquired at $0.93 \times 0.93 \times 3.0$mm resolution.

2.3.3 Image Preprocessing & Tissue Segmentation

For the T_1 map of each subject, we perform B_0-field correction to reduce magnetic field inhomogeniety and B_1-field correction to account for the radio-frequency transmit bias field. For each subject, we rigidly align the corrected T_1 map and T_1w, T_2w, PDw and $FLAIR$ images to the MNI152 1.0 mm nonlinear template, using a two-step registration technique where a second alignment is done after skull-stripping. We apply the N4 inhomogeneity correction algorithm [34] to the T_1w, T_2w, PDw and $FLAIR$ images, and we remove extracerebral voxels using the SPECTRE skull-stripping algorithm [35].

For each study, we use a coarse tissue class segmentation from Topology Preserving Anatomy Driven Segmentation (TOADS) (Bazin and Pham 2008) for HVs and Lesion-TOADS (Shiee et al. 2010) for patients with MS. Both algorithms identify eight normal tissue classes and CSF; Lesion-TOADS also identifies white matter lesions. Since TOADS and Lesion-TOADS employ topological constraints that can cause segmentation errors within the ventricles, which can appear discontinuous on MRI, we correct the ventricular segmentation using the non-topologically constrained maximum membership classes.

To create a brain mask including only the eight normal tissue classes and white matter lesions, we exclude voxels identified as CSF. Furthermore, we exclude any voxels that appear hypointense in the $FLAIR$ image by thresholding the image below the 80th percentile, which has been shown to help correct for CSF segmentation errors [36]. Finally, we exclude any voxels outside the field of view on any image and voxels on the T_1 map with physically implausible values for brain tissue (defined as less than zero or greater than 5s). The voxels that remain after these exclusions comprise the brain mask.

To create a conservative mask of each tissue class, we start with the tissue class segmentation described above and remove any voxels outside of the brain mask. To exclude voxels exhibiting partial volume effects along tissue class borders, we erode the mask of each tissue class using a $3 \times 3 \times 3$ diamond-shaped kernel. Eroded voxels are excluded from model training and validation measures, but are included for whole-image prediction.

2.3.4 QuEEN Map Generation

Within the high-quality dataset, we generate QuEEN maps through leave-one-subject-out cross-validation. That is, for each subject, the QuEEN statistical model is fit on all other subjects in the high-quality dataset, and the resulting model is used to generate the QuEEN map for that subject. For subjects not included in the high-quality dataset, QuEEN maps are generated using the QuEEN statistical model trained on all subjects in the high-quality dataset. The results reported below are based on the high-quality dataset; results including the full dataset are shown in Appendix Figures C.3-C.5.

3 Results

Appendix Figure A.1 shows the estimated coefficient curves for the GAM model fit within each tissue class across all subjects in the high-quality dataset. Figure 2 shows, for one randomly selected subject, two axial slices of the T_1 map (Figure 2a), QuEEN map (Figure 2b), and the magnitude of the difference between them. The difference image is shown on the same scale as the images (Figure 2c) and again on a different scale to show greater detail (Figure 2d). The T_1 and QuEEN maps appear very similar; however, the QuEEN map appears smoother and contains less noise on the border between brain tissue and cerebral spinal fluid (CSF). The subject displayed is an RRMS patient; example images from subjects with PPMS and SPMS and a healthy control subject are shown in Appendix Figure B.2.

Figure 3[1] shows boxplots over subjects of the QuEEN estimation error (shown in orange), QuEEN prediction error (shown in light blue), and T_1 prediction error (shown in dark blue), in terms of root mean squared error (rMSE) within each image and tissue class. The error rates of QuEEN and T_1 maps appear quite similar overall, with differences within certain tissue classes. In particular, the QuEEN maps tend to be more accurate in deep gray matter structures, cerebellar white matter (CBWM) and the brainstem, while the T_1 maps tend to be more accurate in CGM, NAWM and lesions.

Figures 4 and 5 show the results of tests of group differences between disease types. Figure 4 shows the median T_1 in NAWM as estimated by T_1 maps (Figure 4a) and QuEEN maps (Figure 4b). For each pair of groups, the p-value reported corresponds to a one-sided Wilcoxon test that the group on the left has NAWM T_1 greater than the group on the right. These comparisons were chosen based on previous findings [1]. Surprisingly, while none of the expected differences are significant when T_1 maps are used, we do find significant differences using QuEEN maps. Specifically, SPMS patients show a significant or marginally significant increase in NAWM T_1 compared to RRMS patients ($p = 0.018$) and PPMS patients ($p = 0.082$).

Similarly, Figure 5 shows the median T_1 in lesions as estimated by T_1 maps (Figure 5a) and QuEEN maps (Figure 5b). The p-value reported for each group corresponds to a *two-sided* Wilcoxon test for differences. We perform two-sided tests because we do not have strong a-priori beliefs about differences in lesion T_1 between disease types. Again, while none of the differences are significant when T_1 maps are used, we see increased power using QuEEN maps: PPMS patients show marginally elevated T_1 in lesions compared with RRMS patients ($p = 0.082$) and SPMS patients ($p = 0.092$). This finding is supported by earlier work that identified indicators of increased gliosis and axonal loss in PPMS lesions relative to RRMS lesions [37]. The increase in power we observe in both NAWM and lesions is likely due to smaller within-group variance of T_1 using QuEEN maps, which is reflected by slightly narrower confidence bands and fewer outliers in each group (Figures 4 and 5).

4 Discussion

In this paper, we have proposed QuEEN, a new method to estimate T_1 that only requires the acquisition or availability of four conventional MR images. We have demonstrated the accuracy of QuEEN maps by showing that they have similar prediction error compared with T_1 maps and, in fact, have improved accuracy in deep gray matter structures. Given the emerging recognition of the importance in MS of deep gray matter structures, including the thalamus [38, 39, 40], this may represent an important advancement in the study of MS. Furthermore, we have demonstrated the utility of QuEEN maps, as tests of group comparisons of T_1 in NAWM and lesions using QuEEN maps resulted in similar findings but were more powerful than those same tests performed using T_1 maps.

Key to the success of QuEEN is a novel intensity normalization procedure that uses cerebellar gray matter to estimate the "center" of the distribution of image intensities. We also evaluated several other tissue types for this purpose, including NAWM, CSF, and extra-cerebral soft tissue, and we found CBGM

[1] In Figure 3, data points corresponding to lesion masks containing fewer than 500 voxels and caudate masks containing fewer than 1000 voxels are excluded. Such masks usually represent significant overlap with CSF voxels.

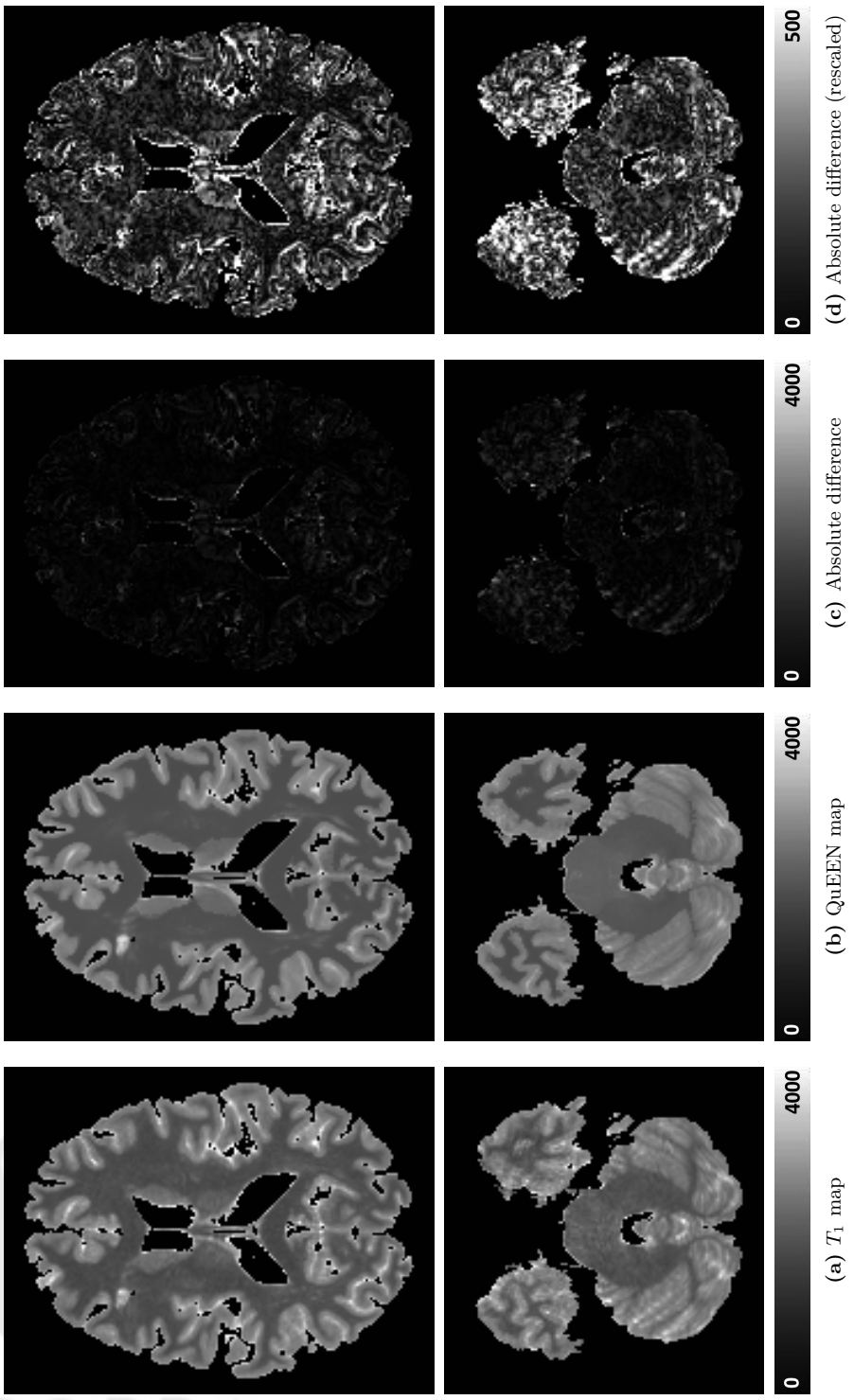


Figure 2: For one randomly selected example subject with RRMS, two axial slices of the T_1 map (a), QuEEN map (b), and the absolute value of the difference between the two on the same scale as the images (c) and rescaled to show greater detail (d).

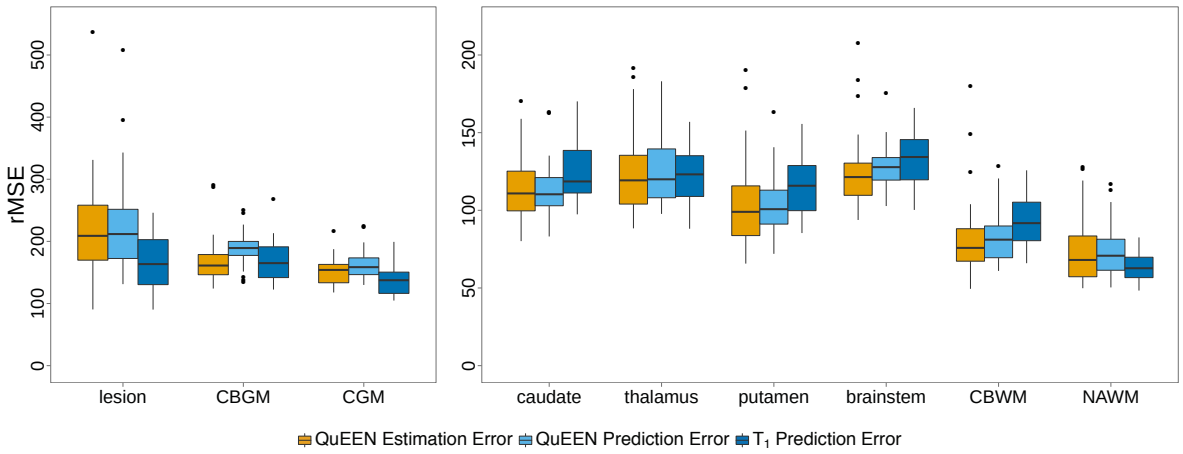
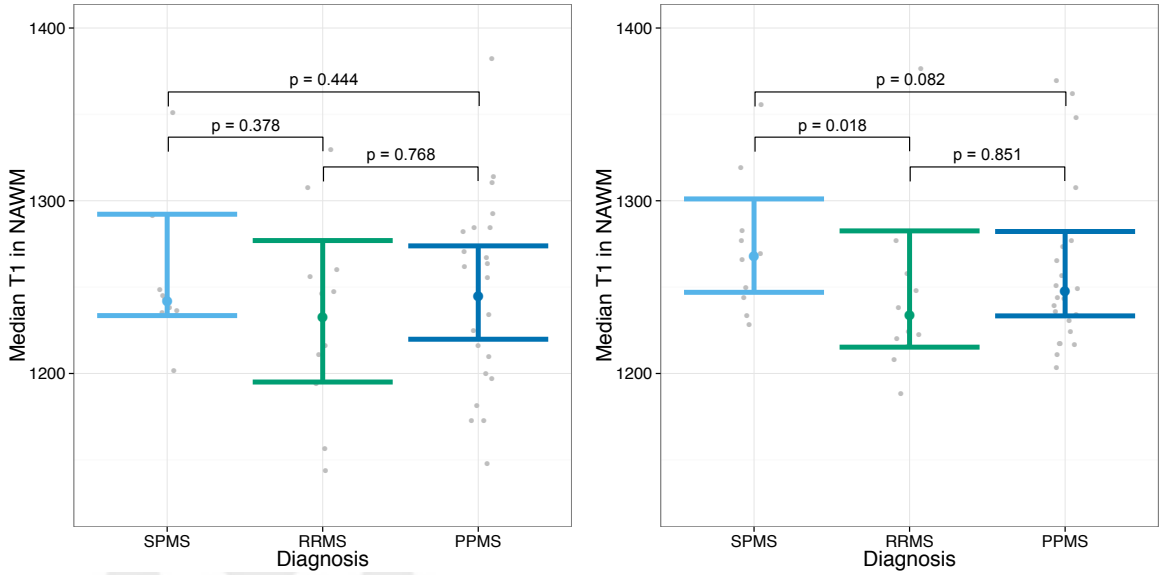
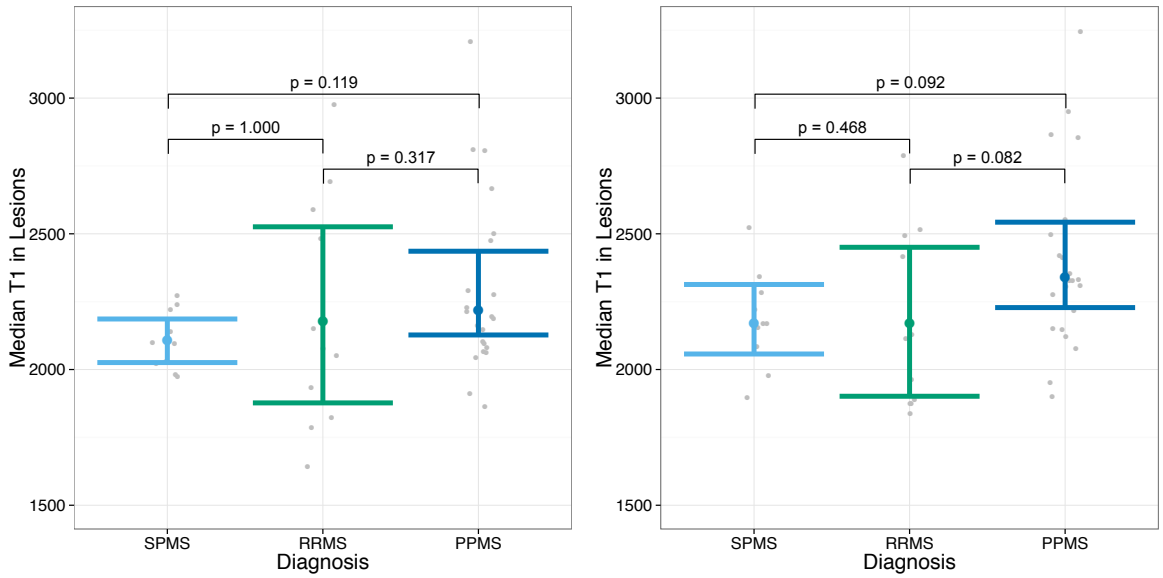


Figure 3: For each tissue class, boxplots of the QuEEN estimation error (shown in orange), QuEEN prediction error (shown in light blue), and T_1 prediction error (shown in dark blue). Each error rate is summarized as the root MSE over all voxels in the eroded mask of each tissue class for a single subject. The boxplots of QuEEN estimation error show the distribution over all subjects in the high-quality dataset, while the boxplots of QuEEN and T_1 prediction errors show the distribution over those subjects in the high-quality dataset who received a second scan. The error rates of QuEEN and T_1 maps appear quite similar overall, with differences within certain tissue classes.



(a) T_1 Maps **(b)** QuEEN Maps

Figure 4: Tests of group differences in median NAWM T_1 using T_1 maps (a) and QuEEN maps (b). Each gray point indicates the median T_1 in NAWM for a single subject. The bars indicate Wilcoxon 95% confidence intervals for the median across subjects in each group. The p-value for each pair of groups corresponds to a *one-sided* Wilcoxon test that the group on the left has NAWM T_1 greater than the group on the right. For example, the test that SPMS patients have NAWM T_1 greater than RRMS patients has $p = 0.378$ using T_1 maps and $p = 0.018$ using QuEEN maps. Due to smaller within-group variance, tests performed using QuEEN maps instead of T_1 maps appear to be more powerful (have smaller p-values) in general.



(a) T_1 Maps **(b)** QuEEN Maps

Figure 5: Tests of group differences in median lesion T_1 using T_1 maps (a) and QuEEN maps (b). Each gray point indicates the median T_1 in lesions for a single subject. The bars indicate Wilcoxon 95% confidence intervals for the median across subjects in each group. The p-value for each pair of groups corresponds to a *two-sided* Wilcoxon test of group differences. While no group differences are significant at $\alpha = 0.05$, there is some evidence that T_1 in lesions is higher for PPMS patients than RRMS or SPMS patients. Tests performed using QuEEN maps instead of T_1 maps again appear to be more powerful.

to have by far the best performance. We believe this is due to the combination of its relatively high similarity across subjects (compared with NAWM) and relatively low noise (compared with CSF and extra-cerebral soft tissue). However, there are some potential issues to consider before employing this normalization approach. First, CBGM can exhibit some changes due to MS disease pathology [41]. From an imaging point of view, these changes may be small in MS, but for patients with other neurological diseases in which CBGM may be affected to a larger degree, the choice of reference region should be revisited before generating QuEEN maps. Second, the cerebellum is close to the receive coil array, and is hence more susceptible to bias-field inhomogeneities. Finally, segmentation of gray and white cerebellar tissue is required. However, this segmentation is not required to be perfect, as the methods we propose are designed to be robust to partial volume effects and minor segmentation errors.

QuEEN also uses a full tissue class segmentation for model training and prediction. The methods currently available for segmentation in diseased populations require tuning and are prone to errors, even when applied by an expert technician. In fact, several of the subjects in our dataset were excluded due to problems with the tissue class segmentation, illustrating the need for more accurate segmentation methods for diseased populations. To account for small errors and partial volume effects, we use eroded tissue class masks for intensity normalization and model training. Alternatively, for the purposes of intensity normalization it may be possible to avoid the tissue class segmentation entirely by using histogram "stripe"-based methods such as that described in [16]. Existing "white stripe" methods could be used to estimate the standard deviation within NAWM; further research would be needed to develop a method to identify the histogram peak location within CBGM.

A limitation of our validation is that we have tested our method on a single protocol acquired on a Siemens scanner and at a single field strength. Additional validation should be conducted on a large, multi-center dataset to assess the site-to-site and scanner-to-scanner reproducibility of QuEEN maps. However, as QuEEN maps do not depend on the careful scanner calibration required for acquisition of traditional T_1 maps, we expect that they may be more reproducible across sites and protocols compared with traditional T_1 maps, which are sensitive to many technical factors.

One advantage of the QuEEN model is that it treats each individual voxel as a separate, independent observation and hence does not require spatial normalization across subjects. As diseased brains can be difficult to normalize to a template [42], avoiding this step is a strength of our approach. While the QuEEN regression residuals may in truth exhibit some spatial correlation, this does not affect the validity of our results, as we rely on out-of-sample prediction rather than model coefficients and their standard errors to evaluate model fit.

QuEEN is a flexible model framework that can be adapted to various contexts through model retraining. For example, while the model specified in this paper utilizes four standard MR images, we also tested the model using only T_1w plus one or more additional predictor images and observed a relatively minor loss of accuracy (results not reported). The QuEEN model can be re-trained to adapt to a setting where the types of images or the protocol under which they were acquired are different from what we have described. For example, the model may need to be retrained to accommodate an FSPGR (Fast SPoiled Gradient Echo) rather than an MPRAGE T_1w image. The QuEEN framework can also be extended to other quantitative modalities.

Finally, QuEEN maps may also be useful for the study of diseases other than MS where T_1 is an important biomarker, including Parkinsons disease, cancers, and diseases of the heart, lungs and abdominal system. The intensity normalization procedure is key to accuracy of the QuEEN model, and should be adapted to each context by identifying a region that is minimally affected by the disease in question and is relatively well imaged. Future research should focus on assessing accuracy of QuEEN maps for different patient populations.

5 Conclusion

In this paper, we have proposed a new way to create T_1 maps retroactively that only requires the acquisition of four standard clinical MR images. The "QuEEN maps" produced using these methods were shown to have similar accuracy and reproducibility compared with traditional acquired T_1 maps. Furthermore, QuEEN maps offer several advantages over traditional T_1 maps. Of primary importance is convenience and *retroactive* availability: QuEEN maps can be computed without any additional scan time using four conventional images. Since these images have historically been included in most clinical and research protocols, our method has the potential to greatly increase the availability of T_1 maps for clinical and research use. Furthermore, QuEEN maps were shown to be more powerful than traditional T_1 maps for tests of differences in T_1 between MS subtypes in NAWM and lesions. Further research is needed to validate QuEEN across scanners, protocols and populations.

References

[1] Vrenken, H. *et al.* Whole-brain t1 mapping in multiple sclerosis: global changes of normal-appearing gray and white matter. *Radiology* **240**, 811–820 (2006).

[2] Larsson, H. *et al.* Assessment of demyelination, edema, and gliosis by in vivo determination of t1 and t2 in the brain of patients with acute attack of multiple sclerosis. *Magnetic resonance in medicine* **11**, 337–348 (1989).

[3] Tofts, P. & Du Boulay, E. Towards quantitative measurements of relaxation times and other parameters in the brain. *Neuroradiology* **32**, 407–415 (1990).

[4] Griffin, C. M. *et al.* The relationship between lesion and normal appearing brain tissue abnormalities in early relapsing remitting multiple sclerosis. *Journal of neurology* **249**, 193–199 (2002).

[5] Vaithianathar, L., Tench, C. R., Morgan, P. S., Lin, X. & Blumhardt, L. D. White matter t1 relaxation time histograms and cerebral atrophy in multiple sclerosis. *Journal of the neurological sciences* **197**, 45–50 (2002).

[6] Parry, A. *et al.* White matter and lesion t1 relaxation times increase in parallel and correlate with disability in multiple sclerosis. *Journal of neurology* **249**, 1279–1286 (2002).

[7] Castriota-Scanderbeg, A., Fasano, F., Filippi, M. & Caltagirone, C. T1 relaxation maps allow differentiation between pathologic tissue subsets in relapsing-remitting and secondary progressive multiple sclerosis. *Multiple sclerosis* **10**, 556–561 (2004).

[8] Parry, A. *et al.* Mri brain t1 relaxation time changes in ms patients increase over time in both the white matter and the cortex. *Journal of Neuroimaging* **13**, 234–239 (2003).

[9] Davies, G. R. *et al.* Normal-appearing grey and white matter t1 abnormality in early relapsing-remitting multiple sclerosis: a longitudinal study. *Multiple Sclerosis* **13**, 169–177 (2007).

[10] Manfredonia, F. *et al.* Normal-appearing brain t1 relaxation time predicts disability in early primary progressive multiple sclerosis. *Archives of neurology* **64**, 411 (2007).

[11] Papadopoulos, K. *et al.* Ti-relaxation time changes over five years in relapsing-remitting multiple sclerosis. *Multiple Sclerosis* **16**, 427–433 (2010).

[12] Neema, M. *et al.* T1-and t2-based mri measures of diffuse gray matter and white matter damage in patients with multiple sclerosis. *Journal of Neuroimaging* **17**, 16S–21S (2007).

[13] Niepel, G. *et al.* Deep gray matter and fatigue in ms. *Journal of neurology* **253**, 896–902 (2006).

[14] Ratchford, J. N. *et al.* Decreased microglial activation in ms patients treated with glatiramer acetate. *Journal of neurology* **259**, 1199–1205 (2012).

[15] Kropholler, M. A. *et al.* Evaluation of reference regions for (r)[^{11}c]pk11195 studies in alzheimer's disease and mild cognitive impairment. *Journal of Cerebral Blood Flow & Metabolism* **27**, 1965–1974 (2007).

[16] Shinohara, R. T. *et al.* Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical* (2014).

[17] Shinohara, R. T., Crainiceanu, C. M., Caffo, B. S., Gaitán, M. I. & Reich, D. S. Population-wide principal component-based quantification of blood-brain-barrier dynamics in multiple sclerosis. *NeuroImage* **57**, 1430–1446 (2011).

[18] Pujol, J. *et al.* Biological significance of iron-related magnetic resonance imaging changes in the brain. *Archives of neurology* **49**, 711 (1992).

[19] Van Waesberghe, J. *et al.* Patterns of lesion development in multiple sclerosis: longitudinal observations with t1-weighted spin-echo and magnetization transfer mr. *American journal of neuroradiology* **19**, 675–683 (1998).

[20] Bakshi, R. *et al.* T2 hypointensity in the deep gray matter of patients with multiple sclerosis: a quantitative magnetic resonance imaging study. *Archives of Neurology* **59**, 62 (2002).

[21] Tjoa, C., Benedict, R., Weinstock-Guttman, B., Fabiano, A. & Bakshi, R. Mri t2 hypointensity of the dentate nucleus is related to ambulatory impairment in multiple sclerosis. *Journal of the neurological sciences* **234**, 17–24 (2005).

[22] Brass, S., Benedict, R. H., Weinstock-Guttman, B., Munschauer, F. & Bakshi, R. Cognitive impairment is associated with subcortical magnetic resonance imaging grey matter t2 hypointensity in multiple sclerosis. *Multiple sclerosis* **12**, 437–444 (2006).

[23] Neema, M. *et al.* Deep gray matter involvement on brain mri scans is associated with clinical progression in multiple sclerosis. *Journal of Neuroimaging* **19**, 3–8 (2009).

[24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2013). URL `http://www.R-project.org/`.

[25] Wood, S. *Generalized additive models: an introduction with R* (CRC press, 2006).

[26] Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 3–36 (2011).

[27] Gu, C. & Wahba, G. Minimizing gcv/gml scores with multiple smoothing parameters via the newton method. *SIAM Journal on Scientific and Statistical Computing* **12**, 383–398 (1991).

[28] Shou, H. *et al.* Shrinkage prediction of seed-voxel brain connectivity using resting state fmri. *NeuroImage* (2014).

[29] Mejia, A. F. *et al.* Improving reliability of subject-level resting-state fmri parcellation with shrinkage estimators. *NeuroImage* (2015).

[30] Miller, D., Johnson, G., Tofts, P., MacManus, D. & McDonald, W. Precise relaxation time measurements of normal-appearing white matter in inflammatory central nervous system disease. *Magnetic resonance in medicine* **11**, 331–336 (1989).

[31] Christensen, K. A., Grant, D. M., Schulman, E. M. & Walling, C. Optimal determination of relaxation times of fourier transform nuclear magnetic resonance. determination of spin-lattice relaxation times in chemically polarized species. *The Journal of Physical Chemistry* **78**, 1971–1977 (1974).

[32] Gupta, R. K. A new look at the method of variable nutation angle for the measurement of spin-lattice relaxation times using fourier transform nmr. *Journal of Magnetic Resonance (1969)* **25**, 231–235 (1977).

[33] Duan, Q., Gelderen, P. & Duyn, J. Improved bloch-siegert based b1 mapping by reducing off-resonance shift. *NMR in Biomedicine* **26**, 1070–1078 (2013).

[34] Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* **29**, 1310–1320 (2010). URL `http://dx.doi.org/10.1109/TMI.2010.2046908`.

[35] Carass, A. *et al.* Simple paradigm for extra-cerebral tissue removal: Algorithm and analysis. *NeuroImage* **56**, 1982–1992 (2011).

[36] Sweeney, E. M. *et al.* Oasis is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in mri. *NeuroImage: clinical* **2**, 402–413 (2013).

[37] Suhy, J. *et al.* 1h mrsi comparison of white matter and lesions in primary progressive and relapsing-remitting ms. *Multiple sclerosis* **6**, 148–155 (2000).

[38] Minagar, A. *et al.* The thalamus and multiple sclerosis modern views on pathologic, imaging, and clinical aspects. *Neurology* **80**, 210–219 (2013).

[39] Zivadinov, R. *et al.* Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting ms during 5 years. *American Journal of Neuroradiology* **34**, 1931–1939 (2013).

[40] Haider, L. *et al.* Multiple sclerosis deep grey matter: the relation between demyelination, neurodegeneration, inflammation and iron. *Journal of Neurology, Neurosurgery & Psychiatry* jnnp–2014 (2014).

[41] Howell, O. W. *et al.* Extensive grey matter pathology in the cerebellum in multiple sclerosis is linked to inflammation in the subarachnoid space. *Neuropathology and applied neurobiology* (2014).

[42] Eloyan, A. *et al.* Health effects of lesion localization in multiple sclerosis: Spatial registration and confounding adjustment. *PloS one* **9**, e107263 (2014).

Funding Acknowledgments

Mejia, Sweeney, and Shinohara are partially funded by the NIH grant RO1 NS085211 from the National Institute of Neurological Disorders and Stroke (NINDS). Mejia is partially funded by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1232825. The study was supported in part by the Intramural Research Program of NINDS. We acknowledge the contribution of the NINDS Neuroimmunology Clinic and the NIH Functional MRI Facility. This work represents the opinions of the researchers and not necessarily that of the granting organizations.

Author Contribution Statement

All authors contributed extensively to the work presented in this paper.

A Coefficient curves from GAM regression models

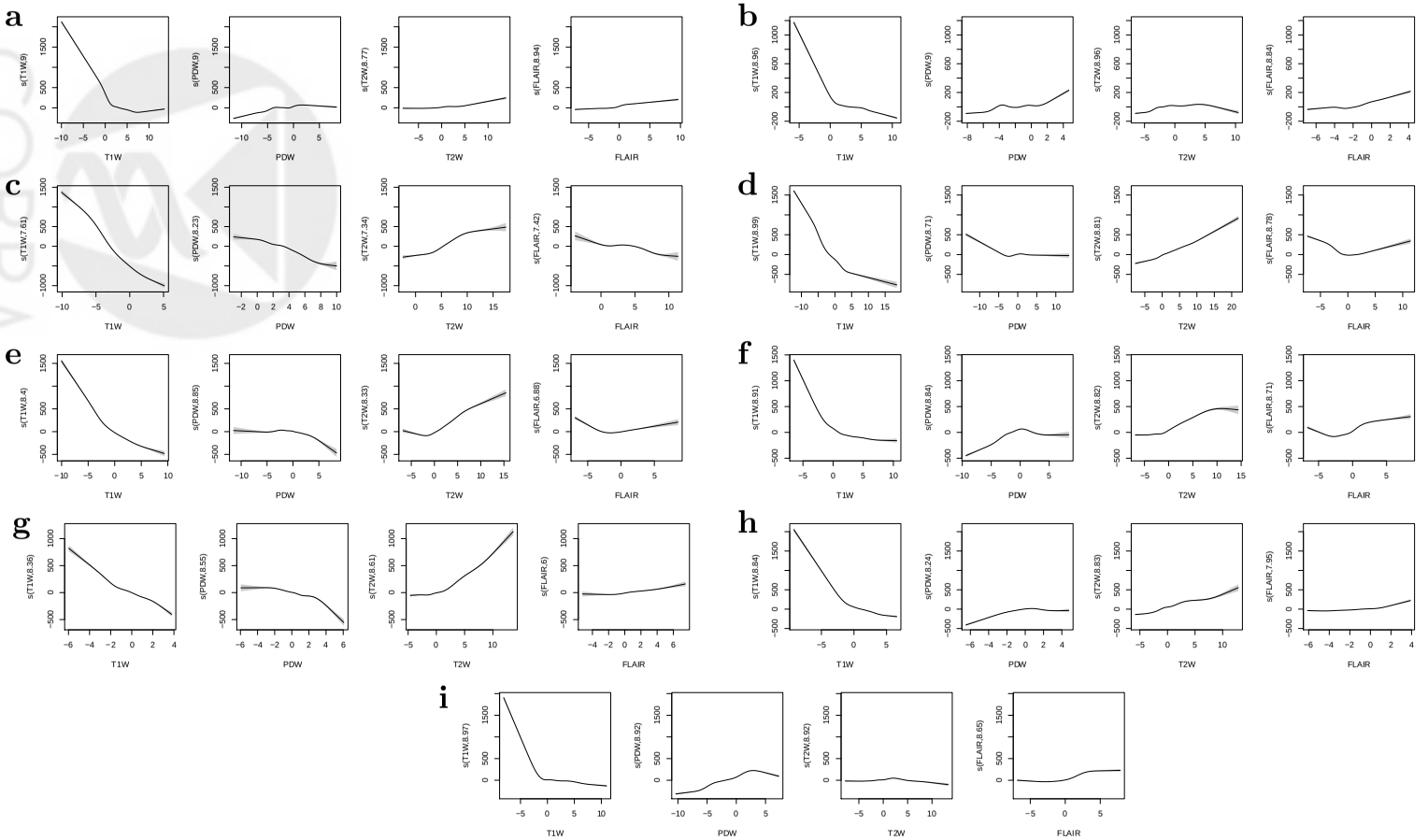


Figure A.1: Estimated coefficient curves for the GAM (spline) regression model within each tissue class: (a) cerebral NAWM, (b) cerebellar white matter, (c) lesion, (d) cortical gray matter, (e) cerebellar gray matter, (f) thalamus, (g) caudate, (h) putamen, and (i) brainstem. For each tissue class, there are four estimated coefficient curves, corresponding to the four predictor sequences in the model. Pointwise 95% confidence bands are shown in gray. For a given tissue class and image, the value of the coefficient function evaluated at a particular (normalized) image intensity is the amount (in ms) that a voxel in that class with that intensity contributes to the predicted T_1 value. For example, (a) shows that a voxel in NAWM with normalized T_1w intensity of -5 would contribute approximately 1300 ms to the predicted T_1 value. The y-axis label indicates the estimated degree of each curve (e.g. degree of 2 indicates a quadratic fit).

B Example images and difference images

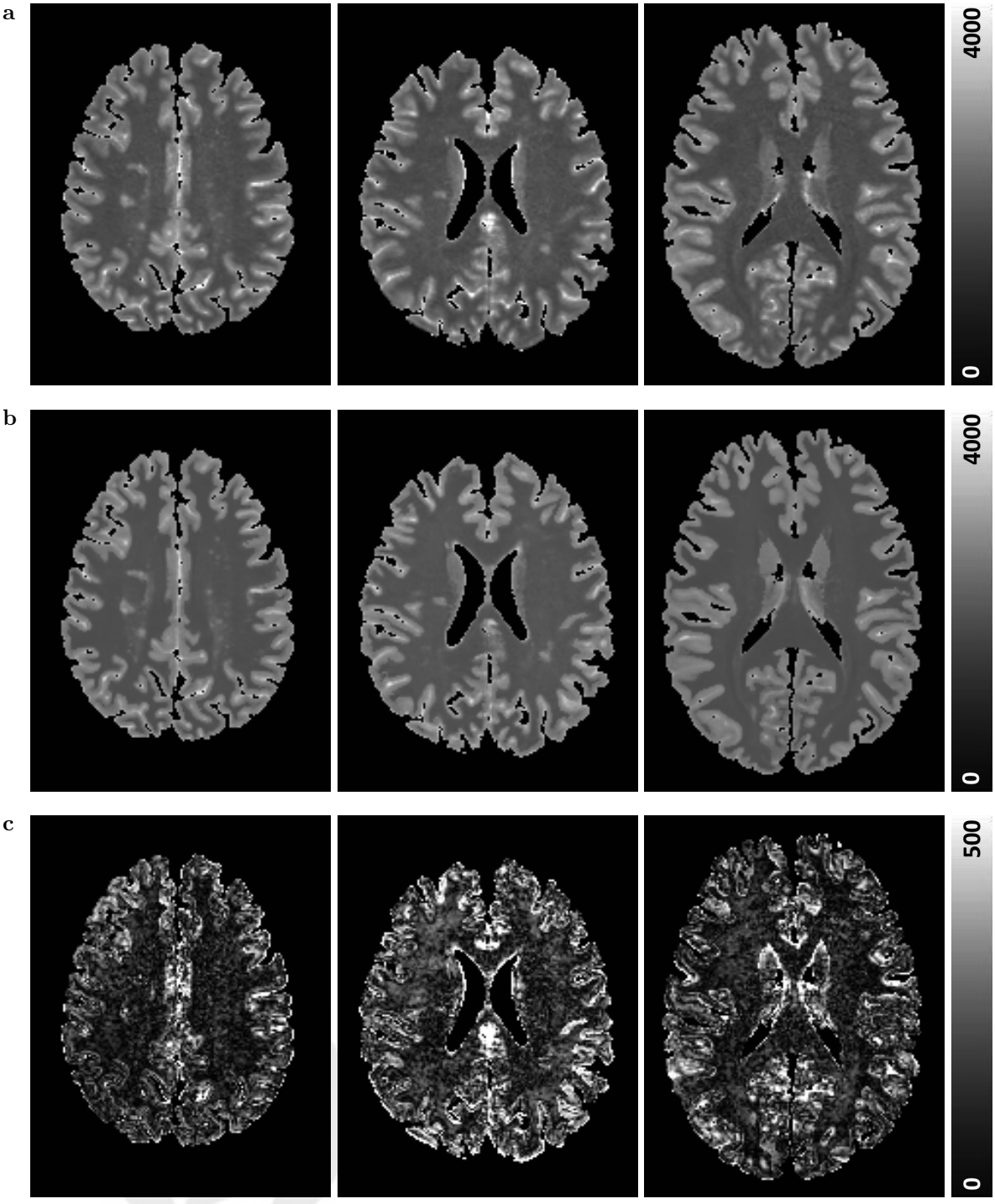


Figure B.2: For three randomly selected example subjects, two axial slices of the T_1 map (a), the QuEEN map (b), and the absolute value of the difference between the two on a different scale (c). The first column shows a subject with SPMS, the middle column shows a subject with PPMS, and the third column shows a healthy volunteer.

C Results on full validation set

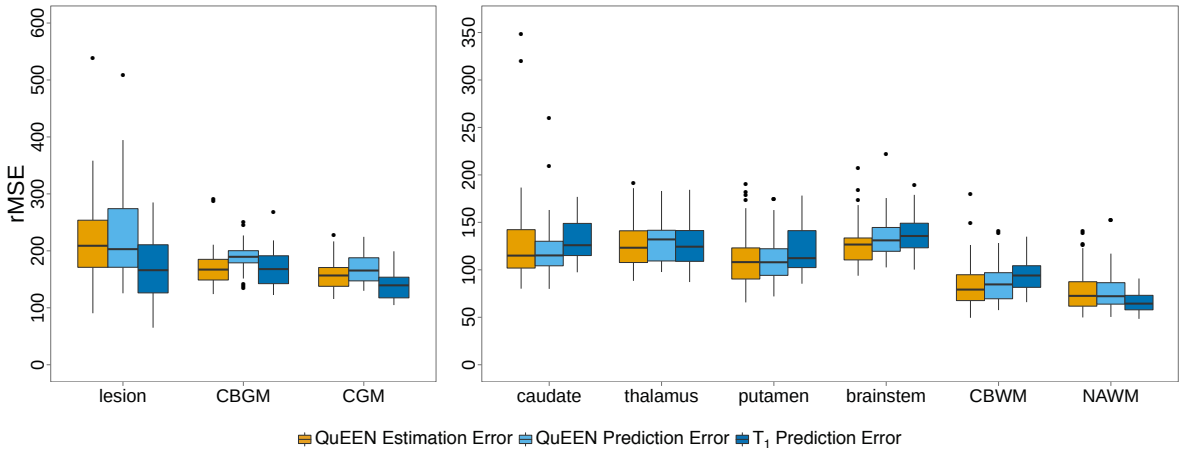


Figure C.3: For each tissue class, boxplots of the QuEEN estimation error (shown in orange), QuEEN prediction error (shown in light blue), and T_1 prediction error (shown in dark blue), using the full dataset. Each error rate is summarized as the root MSE over all voxels in the eroded mask of each tissue class for a single subject. The boxplots of QuEEN estimation error show the distribution over all subjects, while the boxplots of QuEEN and T_1 prediction error show the distribution over those subjects who received a second scan. We see more outliers compared with Figure 3, but the relationship between the error rates in each tissue class remains relatively unchanged.

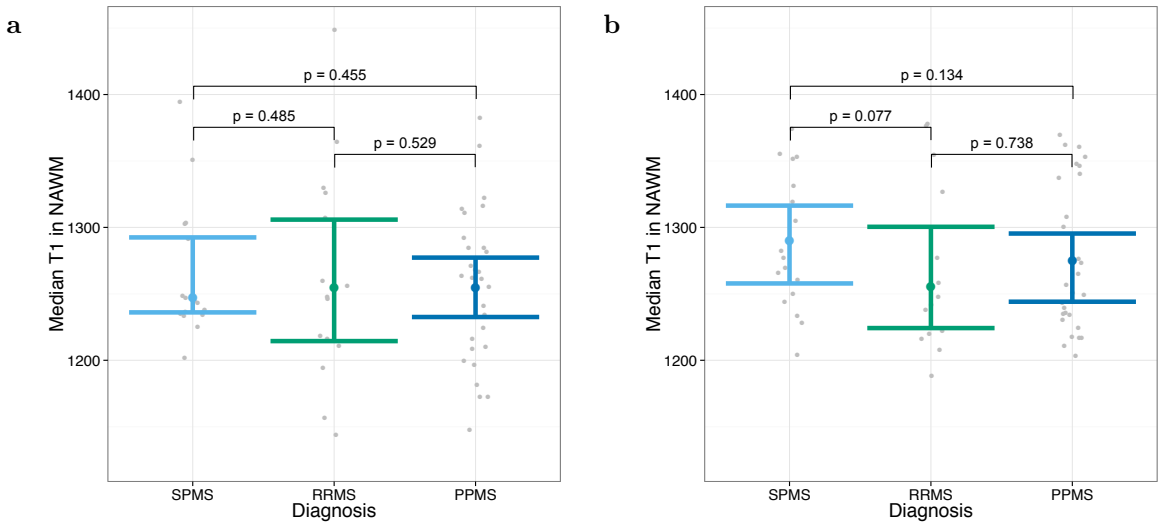


Figure C.4: Tests of group differences in median NAWM T_1 using T_1 maps (a) and QuEEN maps (b), using the full dataset. Each gray point indicates the median estimate of T_1 in NAWM for a single subject. The bars indicate Wilcoxon 95% confidence intervals for the median across subjects in each group. The p-value for each pair of groups corresponds to a one-sided Wilcoxon test that the group on the left has NAWM T_1 greater than the group on the right. For example, the test that SPMS patients have NAWM T_1 greater than RRMS patients has $p = 0.485$ using the T_1 maps and $p = 0.077$ using the QuEEN maps. As seen in Figure 4, tests performed using QuEEN maps instead of T_1 maps appear to be more powerful.

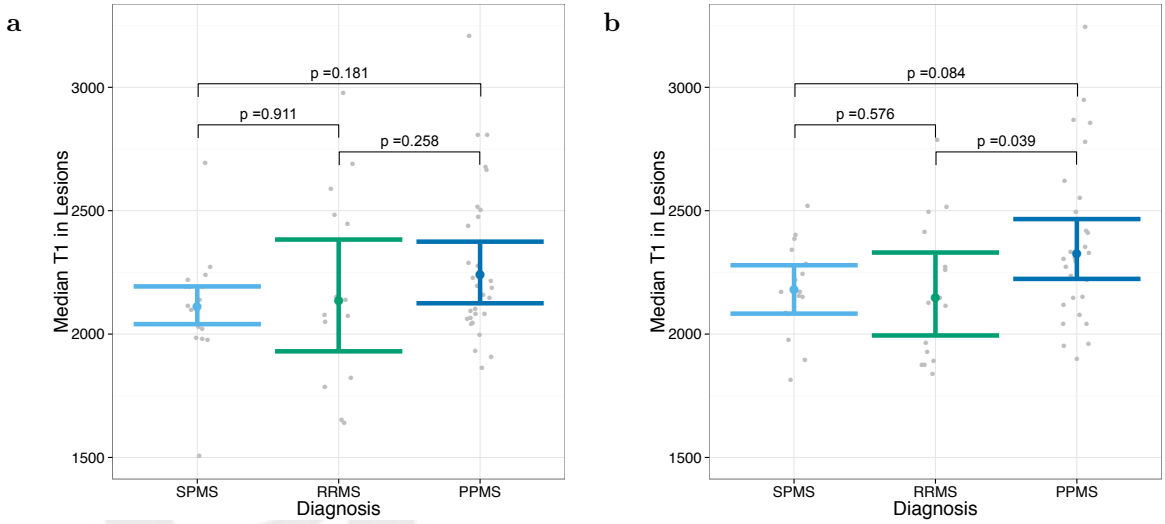


Figure C.5: Tests of group differences in median T_1 in lesions using T_1 maps (a) and QuEEN maps (b), using the full dataset. Each gray point indicates the median estimate of T_1 in lesions for a single subject. The bars indicate Wilcoxon 95% confidence intervals for the median across subjects in each group. The p-value for each pair of groups corresponds to a *two-sided* Wilcoxon test of group differences. While no group differences are significant using T_1 maps, two differences are significant or marginally significant using QuEEN maps: PPMS patients have significantly elevated T_1 in lesions compared with RRMS patients ($p = 0.039$) and marginally elevated T_1 compared with SPMS patients ($p = 0.084$). Again, tests performed using QuEEN maps instead of T_1 maps appear to be more powerful.

D Tests of group differences with healthy controls

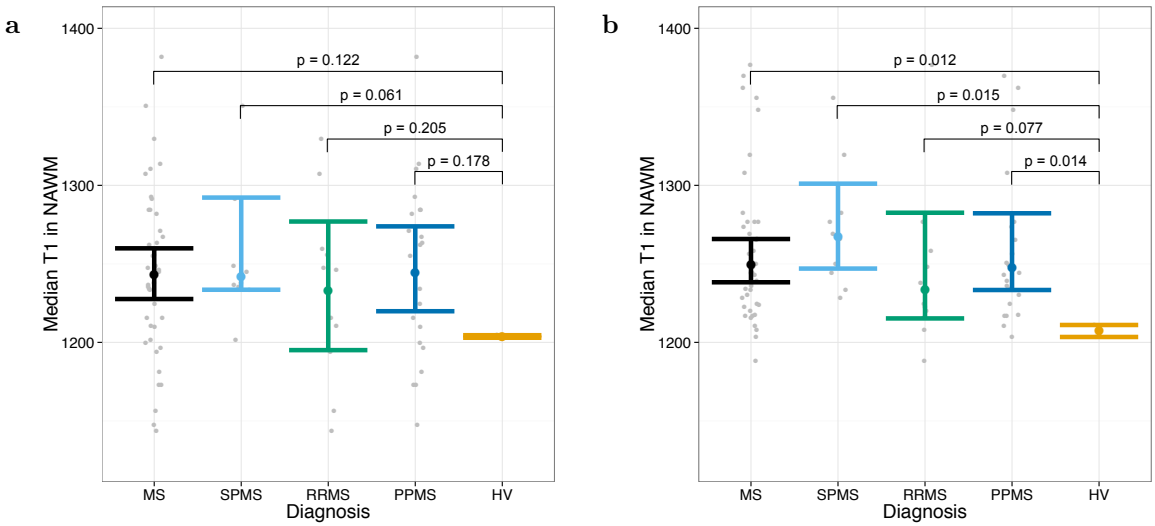


Figure D.6: Exploratory tests of group differences in median NAWM T_1 between MS groups and healthy volunteers using T_1 maps (a) and QuEEN maps (b), using the high-quality dataset. The bars indicate the Wilcoxon 95% confidence intervals for the median across subjects in each MS group and the 50% confidence interval for the HV group. The p-value for each pair of groups corresponds to a one-sided Wilcoxon test that the group on the left has NAWM T_1 greater than the HV group. Using T_1 maps, none of the MS groups show significantly elevated T_1 relative to HVs, though the SPMS group shows a marginally significant elevation ($p = 0.061$). However, using QuEEN maps, the MS, SPMS and PPMS groups show significantly elevated T_1 relative to HVs ($p = 0.012$, $p = 0.015$ and $p = 0.014$, respectively), and the RRMS group shows a marginally significant elevation ($p = 0.077$). The p-value for all groups is smaller when QuEEN maps rather than T_1 maps are used. This provides further evidence that using QuEEN maps results in more powerful tests of group differences.