Subcortical Gray Matter Segmentation and Voxel-Based Analysis Using Transverse Relaxation and Quantitative Susceptibility Mapping With Application to Multiple Sclerosis

Dana Cobzas, PhD,[1,2] Hongfu Sun, BSc,[1] Andrew J. Walsh, MD, PhD,[1] R. Marc Lebel, PhD,[1] Gregg Blevins, MD,[3] and Alan H. Wilman PhD[1]*

Purpose: To investigate subcortical gray matter segmentation using transverse relaxation rate (R_2*) and quantitative susceptibility mapping (QSM) and apply it to voxel-based analysis in multiple sclerosis (MS).

Materials and Methods: Voxel-based variation in R_2* and QSM within deep gray matter was examined and compared to standard whole-structure analysis using 37 MS subjects and 37 matched controls. Deep gray matter nuclei (caudate, putamen, globus pallidus, and thalamus) were automatically segmented and morphed onto a custom atlas based on QSM and standard T_1-weighted images. Segmentation accuracy and scan–rescan reliability were tested.

Results: When considering only significant regions as returned by the multivariate voxel-based analysis, increased R_2* and QSM was found in MS subjects compared to controls in portions of all four nuclei studied ($P < 0.002$). For R_2*, regional analysis yielded at least 66-fold improved P-value significance in all nuclei over standard whole-structure analysis, while for QSM only thalamus benefited, with 5-fold improvement in significance. Improved segmentation over standard methods, particularly for globus pallidus (2.8 times higher Dice score), was achieved by incorporating high-contrast QSM into the atlas. Voxel-based reliability was highest for QSM (<1% variation).

Conclusion: Automatic segmentation of iron-rich deep gray matter can be improved by incorporating QSM. Voxel-based evaluation yielded increased R_2* and QSM in MS subjects in all four nuclei studied with R_2*, benefiting the most from localized analysis over whole-structure measures.

Iron-sensitive quantitative magnetic resonance imaging (MRI) has shown promise for evaluating iron-rich deep gray matter (GM) in the brain, where ferric iron stored in ferritin is a strong source of image contrast.[1–3] Transverse relaxation rates R_2 and R_2* have been used, as well as phase imaging and quantitative susceptibility mapping (QSM), which is derived from phase imaging but eliminates the field orientation dependency and nonlocal effects.[3] Postmortem studies have validated QSM and R_2* methods for use in deep GM by demonstrating high correlations with iron content.[4–8] One application of these methods has been in multiple sclerosis (MS), which is a demyelinating disease that also has a neurodegenerative component that affects GM. For example, subcortical GM changes in R_2, R_2*, phase, and QSM have been demonstrated in MS that suggest increased iron accumulation compared to healthy controls.[9–12] In addition, R_2* mapping has also been used to examine longitudinal subcortical GM changes in MS.[13] These MS studies have focused on regions of particularly high iron content, including the large structures of the basal ganglia (globus pallidus, putamen, and caudate) and the thalamus. However, the QSM and R_2* subcortical GM

TABLE 1. Subject Demographic Information			
Demographics	Patients ($N = 37$)	Controls ($N = 37$)	P-value
Gender (M/F)	6/31	6/31	1.00
Age [y] mean [range]	35.64 [19.5–51.4]	35.70 [21.7–54.5]	0.97
Disease duration [y] mean [range]	5.28 [0.7–10.7]	—	
EDSS mean [range]	2.35 [1–5]	—	
MSSS mean [range]	4.56 [0.9–9.2]	—	

studies have generally analyzed large regions-of-interest (ROIs) that include the whole-structure, which does not supply precise location of iron accumulation. Within a structure, these large ROIs may lead to averaging of discrete regions of high iron accumulation with neighboring areas of no change (or loss), thus masking local changes. Localized assessment of brain iron within deep GM nuclei is also of interest to examine both anatomically and functionally distinct sections that may be biochemically distinct and have differences in iron metabolism.

A recent QSM and R_2* study by Rudko et al.[14] performed voxel-based analysis on MS subjects and healthy controls at 7.0T using a standard whole brain registration approach based on only T_1-weighted (T_1w) images. This work found voxel-based changes in deep GM between MS subjects and controls with QSM, but not with R_2*. However, when using high-field imaging (3.0T or greater), standard brain registration and segmentation methods may perform poorly for some deep GM nuclei due to low T_1w contrast,[15] particularly for the globus pallidus and thalamus, which provide reduced subcortical GM boundary discrimination from neighboring white matter. Increasing the magnetic field remains beneficial to obtain increased iron sensitivity when using susceptibility-based methods,[1,16,17] but reduced T_1w contrast may affect the quality of subcortical segmentation. Since R_2* and QSM provide high contrast of iron-rich tissue, they may aid segmentation of deep GM. In particular, a QSM atlas has recently been developed at 3T to provide improved deep GM segmentation.[18] The purpose of our work was to investigate subcortical GM segmentation using R_2* and QSM combined with standard T_1w images, and then to perform voxel-based analysis using R_2* and QSM in healthy subjects and subjects with MS.

Materials and Methods

Overview

We examined the value of an R_2* or QSM deep GM atlas combined with standard T_1w images at 4.7T, following a multiatlas approach.[19] We tested the precision of this segmentation scheme against the standards of manual segmentation and FSL FIRST,[20] and examined scan–rescan reliability of voxel-based QSM and R_2*

measurements. Voxel-based differences in QSM and R_2* between relapsing-remitting MS (RRMS) and healthy controls were then investigated to identify focal abnormalities within four iron-rich regions: the globus pallidus, putamen, thalamus, and caudate nucleus.

MRI Methods

Imaging was performed at 4.7T (Varian Inova, Palo Alto, CA). Two imaging sequences were acquired in the same session: 3D T_1w volumetric imaging using magnetization-prepared rapid gradient-echo (MPRAGE) (10° flip, TE/TR 4.5/8.5 msec, inversion time to start of readout 300 msec, sequential phase encoding, 84 slices, 2 mm thick, in-plane 0.9 × 0.9 mm^2, 284 × 222 × 84 matrix, acquisition time 4.8 min) and R_2*/QSM using 3D multiecho gradient echo (10° flip, TE1/TR 2.9/44 msec, 10 echoes, echo spacing 4.1 msec, 80 slices, 2 mm thick, in-plane 1 × 1 mm^2, 256 × 160 × 80 matrix, acquisition time 9.4 min). Both R_2* and QSM were reconstructed from the same multiecho images using previously introduced methods.[9,21] Briefly, R_2* used a 3D linear field gradient correction to compensate for air–tissue susceptibility effects and then a monoexponential fit.[9] For QSM, a field map was estimated from the multiecho data, followed by background field removal using regularization-enabled sophisticated harmonic artifact reduction for phase data (RESHARP)[21] and dipole inversion using total variation regularization.[22–24]

Prior to segmentation, R_2* and QSM from each subject were rigidly aligned with the T_1w images and interpolated to the same resolution. Bias field intensity normalization for T_1w images was performed using Freesurfer.[25]

Subjects

Institutional ethical approval and informed consent was obtained from each subject prior to the study. Thirty-seven subjects with RRMS and 37 age- and gender-matched controls were recruited. Patients were enrolled using the following inclusion criteria: a diagnosis of RRMS according to the McDonald criteria[26] and an Expanded Disability Status Scale (EDSS) score less than or equal to 5.0. Exclusion criteria for all subjects were: non-MS-related neurological disease, significant medical illness, and MRI contraindications. Baseline characteristics were compared between groups using one-way analysis of variance (ANOVA) for age and a nonparametric test for sex. There were no significant sex ($P = 1$) or age ($P = 0.97$) differences between patients and controls (Table 1). For each patient, the EDSS was measured by a neurologist (G.B).

TABLE 2. Quality of Subcortical Segmentation Using Comparative Dice Scores[a]			
Region	FSL FIRST T_1-weighted	Multi-atlas T_1w+ R_2*	Multiatlas T_1w+QSM
Thalamus	78 ± 6	86 ± 1	87 ± 1
Caudate	74 ± 1	79 ± 4	82 ± 1
Putamen	82 ± 2	87 ± 1	87 ± 1
Globus pallidus	29 ± 10	82 ± 3	83 ± 3

[a]Score of 100 is perfect agreement with manual segmentations. Mean ± standard deviation reported.

Image Alignment and Segmentation

AUTOMATIC SEGMENTATION COMPARISON. Many automatic methods for subcortical segmentation have been developed[27] and are publicly available (eg, FSL FIRST[20] and FreeSurfer[25]). We implemented a multiatlas segmentation method[19] using internal atlases based on both T_1w images and either R_2* or QSM and compared to manual segmentation and to a public subcortical GM segmentation tool: FSL FIRST distributed as part of the FSL package.[20] Although other methods such as FreeSurfer are available, we used FSL FIRST, since it has seen common use in application to iron-rich deep GM in MS (eg, 11,12,14) and has demonstrated slightly improved performance over other methods for this application.[15] The segmentation comparison evaluated the overlap between manual and the automatic methods using cross-validation and comparative Dice scores, which measure the extent of spatial overlap between the manual ground truth and automatic segmentation.

MANUAL SEGMENTATIONS. The putamen, globus pallidus, caudate, and thalamus were manually segmented by one expert (A.J.W.) in four healthy subjects using ImageJ software.[28] These four subjects were a subset of the healthy controls studied with ages 30, 30, 40, 49 years. To compensate for intrasubject motion between scans, data from T_1w images and R_2*/QSM were previously rigidly aligned based on mutual information using the Symmetric Normalization (SyN) automatic registration method in the Advanced Normalization Tools (ANTS) package.[29] The manual segmentation was performed using the single modality with highest contrast, except for the thalamus, which used both T_1w and R_2*/QSM by viewing both sets of images simultaneously. Specifically, the caudate and putamen were segmented on T_1w only, the globus pallidus on R_2*/QSM data only, and the thalamus on both T_1w and R_2*/QSM. R_2* and QSM were considered independently, giving rise to two manual atlases in each case (T_1w + R_2*, and T_1w+ QSM).

SUBCORTICAL SEGMENTATION AND ATLAS ALIGNMENT. A series of steps were performed based on existing publicly available software. Following the method of Heckemann et al,[19] to achieve the multiatlas alignment the four manually segmented images (atlases) were propagated to each subject image using automatic nonlinear registration performed using SyN.[29] The labels were propagated using the same image types as the manual segmentation

described above. The registered labels were fused using simultaneous truth and performance level estimation (STAPLE)[30] to produce the optimal segmentation of each subject image.

The segmentation data from all subjects was next registered with a shape atlas that provides the common space for voxel-based statistical analysis. The shape atlas was chosen as one of the manually labeled datasets. No erosion was applied except for the thalamus label, where 1-pixel boundary erosion was used to reduce segmentation and alignment errors. The registration of segmented images with the shape atlas was done using image-based deformable registration SyN[29] applied to the smoothed characteristic images. The characteristic images contain all structures, having one distinct intensity value inside each structure. This calculated nonlinear deformation was applied to the R_2* and QSM images from each subject to bring them into the common atlas space, which was smoothed with a Gaussian filter for statistical analysis.

RELIABILITY OF IMAGE ALIGNMENT METHODS. Scan–rescan tests were performed on five control subjects ~1 hour apart, with the subject exiting the MRI system between runs. The scans were aligned with the shape atlas following the image alignment procedure described above. Percent differences between images were recorded for each structure in two ways using either an absolute voxel-by-voxel difference or a difference between whole-structure averages.

QUANTITATIVE IMAGE CONTRAST MEASUREMENT. Image contrast measurements from T_1w, R_2*, and QSM images were made using the 3D automatic subcortical segmentations from all subjects. The four deep GM structures were measured bilaterally relative to the internal capsule, which is adjacent to the globus pallidus and thalamus. Normalized image contrast is reported as the signal difference between each GM structure and the internal capsule, divided by the absolute signal of the internal capsule. The denominator was the absolute value because the internal capsule can be negative in QSM.

Statistics

Having all data aligned in the shape atlas space, statistical volumetric analysis (voxel-based morphometry, VBM) was performed to evaluate iron accumulation within the four subcortical GM regions. Both R_2* and QSM were studied using the same protocol to examine regional significant differences between RRMS patients and healthy subjects. A general linear regression model dependent

TABLE 3. Mean Normalized Image Contrast[a] of Gray Matter Regions Relative to Internal Capsule

Region	T_1-weighted	R_2*	QSM
Thalamus	-0.15 ± 0.02	0.16 ± 0.09	1.12 ± 0.08
Caudate	-0.31 ± 0.03	0.20 ± 0.10	1.80 ± 0.17
Putamen	-0.20 ± 0.03	0.45 ± 0.13	1.82 ± 0.18
Globus pallidus	-0.006 ± 0.01	1.26 ± 0.25	3.54 ± 0.54

[a]Normalized image contrast is the signal difference normalized by the absolute value of the background signal. (Region – IC)/abs(IC), where IC is Internal Capsule.

on the R_2*/QSM values was employed, with the group as independent variable after covarying for age. Age is included in the regression model since an age-R_2*/QSM correlation is expected in some brain regions.[31,32] SurfStat[33] was used to test this linear model. SurfStat is a MatLab (MathWorks, Natick, MA) toolbox for statistical analysis that uses Random Field Theory (RFT), a recent body of mathematics defining theoretical results for smooth statistical maps.[34] The group effects were determined by testing the significance of the corresponding regression parameters with an uncorrected P-value of 0.05.

As well as detailed statistical maps for R_2*/QSM values, we also investigated group effects on the volume-averaged R_2*/QSM of each whole-structure, averaging left and right sides, to enable comparison to past work that has mainly focused on large ROI analysis.

Results
Subcortical Segmentation Validation
Table 2 quantifies segmentation results for all subjects by reporting the comparative Dice scores (100 = perfect alignment). The multiatlas method which incorporated R_2* or QSM into the segmentation process achieved results closer to manual segmentation than FSL FIRST. In particular for the globus pallidus, the multiatlas methods provided 2.8 times higher Dice scores. The multiatlas method using T_1w with R_2*, or with QSM, gave similar results. Given the higher contrast and smoother edges of QSM, only segmentation with T_1w and QSM was used for the rest of the study.

Table 3 reports mean normalized image contrast from all subjects. In particular, the poor contrast for the globus pallidus with T_1w imaging is evident, with little discrimination from the internal capsule. However, T_1w imaging provides adequate contrast for the caudate and putamen, although poorer results in the thalamus. In contrast, QSM and R_2* provide high contrast differences for globus pallidus. Figure 1 shows images overlaid with segmentation results for a typical healthy subject. The FSL FIRST segmentation tracings (second row) clearly misrepresent the boundaries of the globus pallidus and pulvinar thalamus in comparison to the manual segmentation (bottom row). More effective segmentation is achieved with the combined multiatlas method incorporating T_1w and QSM (third row).

Note that the multiatlas method does not include a final smoothing step, thus the boundaries appear fuzzier than FSL-FIRST, which includes a regularization term to make the boundaries smoother.

Reliability of Image Alignment Methods
Table 4 reports scan–rescan results as percent differences between corresponding aligned scans. Voxel-by-voxel absolute differences had an average error of 5.3% for R_2* and 0.7% for QSM. Bilateral whole-structure differences yield reduced errors of 2.4% for R_2* and 0.4% for QSM. Figure 2 shows the voxel-by-voxel percent differences for one of the subjects. The QSM result is uniformly low, while the R_2* result has regions of substantial variation.

Group Maps of R_2* and QSM
Mean R_2* and QSM in the aligned data, averaged over controls or patients, are shown in Fig. 3 (top row). Similar patterns are observed in the R_2* and QSM group maps. Within each group, there is a high degree of left–right symmetry in the mean values. Using whole-structure measurements, comparisons of left to right sides generally yielded no significance, except for asymmetry in caudate nucleus R_2* (6.5% patients, $P = 0.02$; 4% controls, $P = 0.02$) and globus pallidus in patients only (R_2*: 6.5%, $P = 0.01$; QSM 6.5%, $P = 0.02$). All QSM values are reported as normalized by subtracting the mean internal capsule value, which was similar ($P = 0.304$) in patients (–0.055 ± 0.009 ppm) and controls (–0.053 ± 0.008 ppm).

Maps of the standard deviation divided by the mean are shown in Fig. 3 (middle row). Note that there is substantial within-group variation at some edge points near the boundaries of the thalamus and caudate, particularly for R_2*, where borders are less distinct than QSM. No erosion was used on the atlas labels except for 1-pixel local erosion in the thalamus. Additional erosion of 2 and 4 pixels was tested for all structures and found to not change the significance of the results, indicating robustness of the alignment and segmentation method. Away from the edges, the standard deviation results within groups indicate the variation within each structure between individuals, which tends to
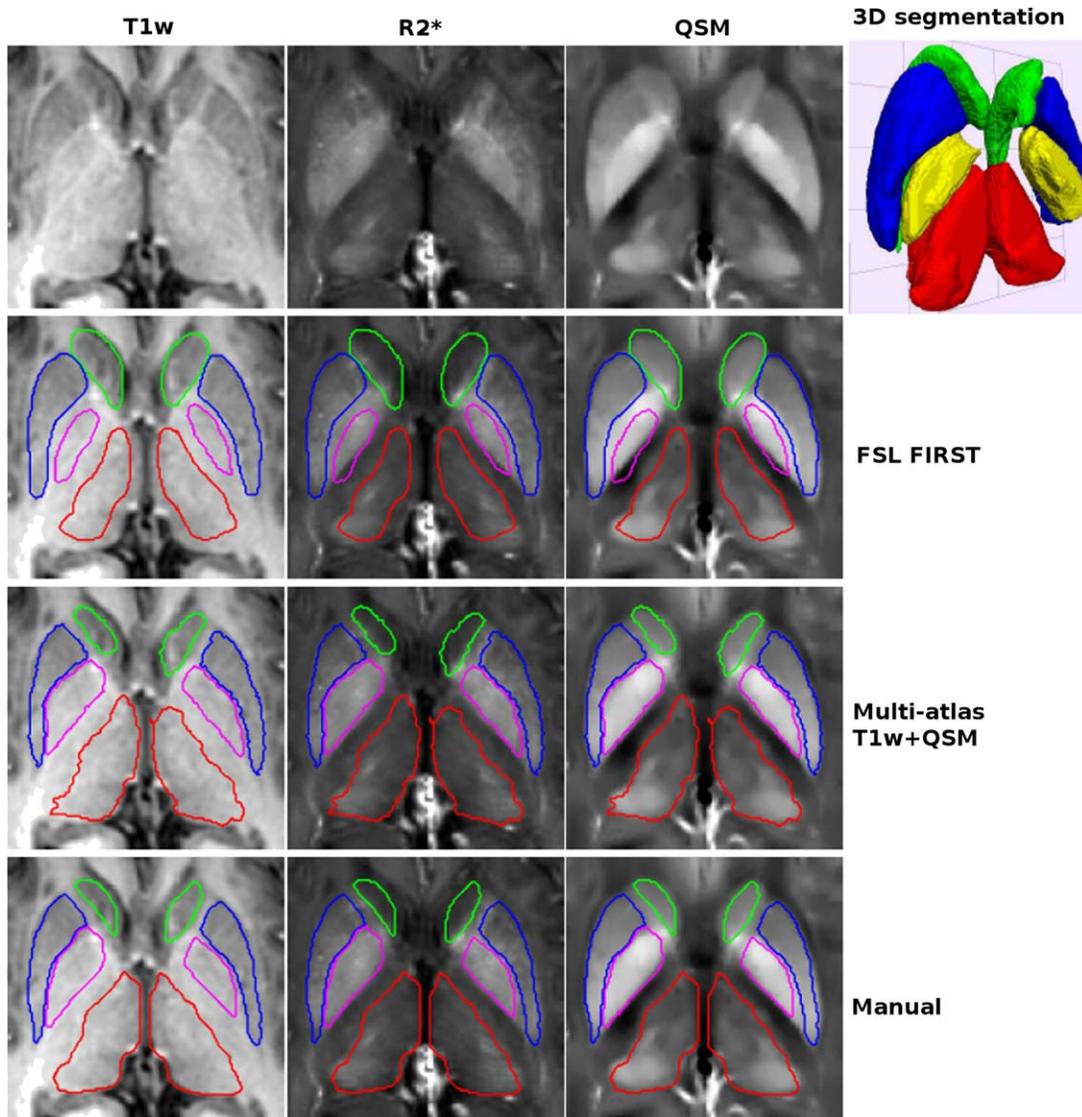
FIGURE 1: Comparison of automatic segmentation methods to manual segmentation in one control subject (female, 31 years). Images from T_1w, R_2*, and QSM are shown in respective columns left to right. Segmentation results are overlaid in respective rows for FSL FIRST, multiatlas alignment using T_1w and QSM, and manual segmentation. The segmentation for this slice is color coded: thalamus (red), caudate (green), putamen (blue), globus pallidus (magenta). A full 3D oblique view of the manual segmentation is shown in the top row at right.

be 10–20% for most pixels in R_2* and 15–25% in QSM. Notably, the anterior portion of the thalamus exhibits far greater variation in QSM.

The positive percent differences in R_2* and QSM, where the voxel had larger values in the group patient map than controls, are shown for the same slice in the bottom

TABLE 4. Mean Percent Differences Between Repeated Scans in Atlas Space From 5 Volunteers

Measure	Thalamus	Caudate	Putamen	Globus pallidus	Mean
mean R_2* voxel[a]	5.8%	5.4%	5.0%	5.2%	5.3%
mean R_2* struct[b]	3.6%	1.6%	1.8%	2.7%	2.4%
mean QSM voxel	0.9%	0.6%	0.7%	0.8%	0.7%
mean QSM struct	0.4%	0.2%	0.4%	0.6%	0.4%

[a]Voxel-by-voxel absolute difference then average per structure.
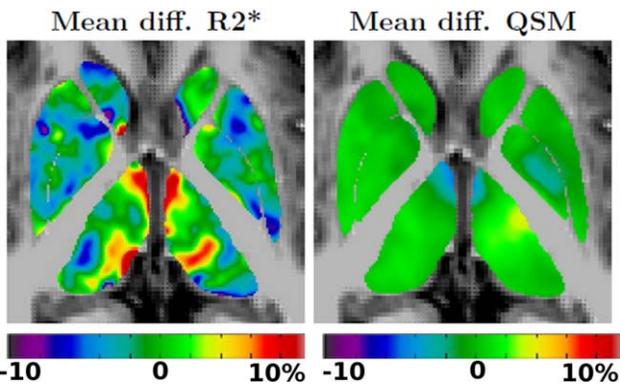[b]Average per structure then difference.

FIGURE 2: Reliability images for one subject. Signed mean percent differences between the baseline and repeated scan are shown for R_2^* (left) and QSM (right).

row of Fig. 3. Voxels with negative difference are left blank and were rare. In this slice shown, a series of negative difference voxels are located along the medial border of the

thalamus, particularly the anterior portion. Across the complete deep GM volume considered, most voxels exhibited positive percent differences with larger mean values in the patient map (99% for QSM and 91% for R_2^*).

Statistical Maps of R_2^* and QSM From Between-Groups Comparisons

In Fig. 4, detailed statistical maps of R_2^* and QSM values for between-groups comparisons are presented for two slices through the 3D volume. Localized differences are illustrated by both the voxel-by-voxel t-statistic and the SurfStat corrected P-clusters ($P < 0.05$). Portions of all deep GM nuclei show significance; however, R_2^* clusters are generally more sporadic and of smaller size. The shape of the significant QSM clusters have a high degree of left–right symmetry. However, there is a left–right asymmetry in the actual P-values obtained for QSM in thalamus and caudate.
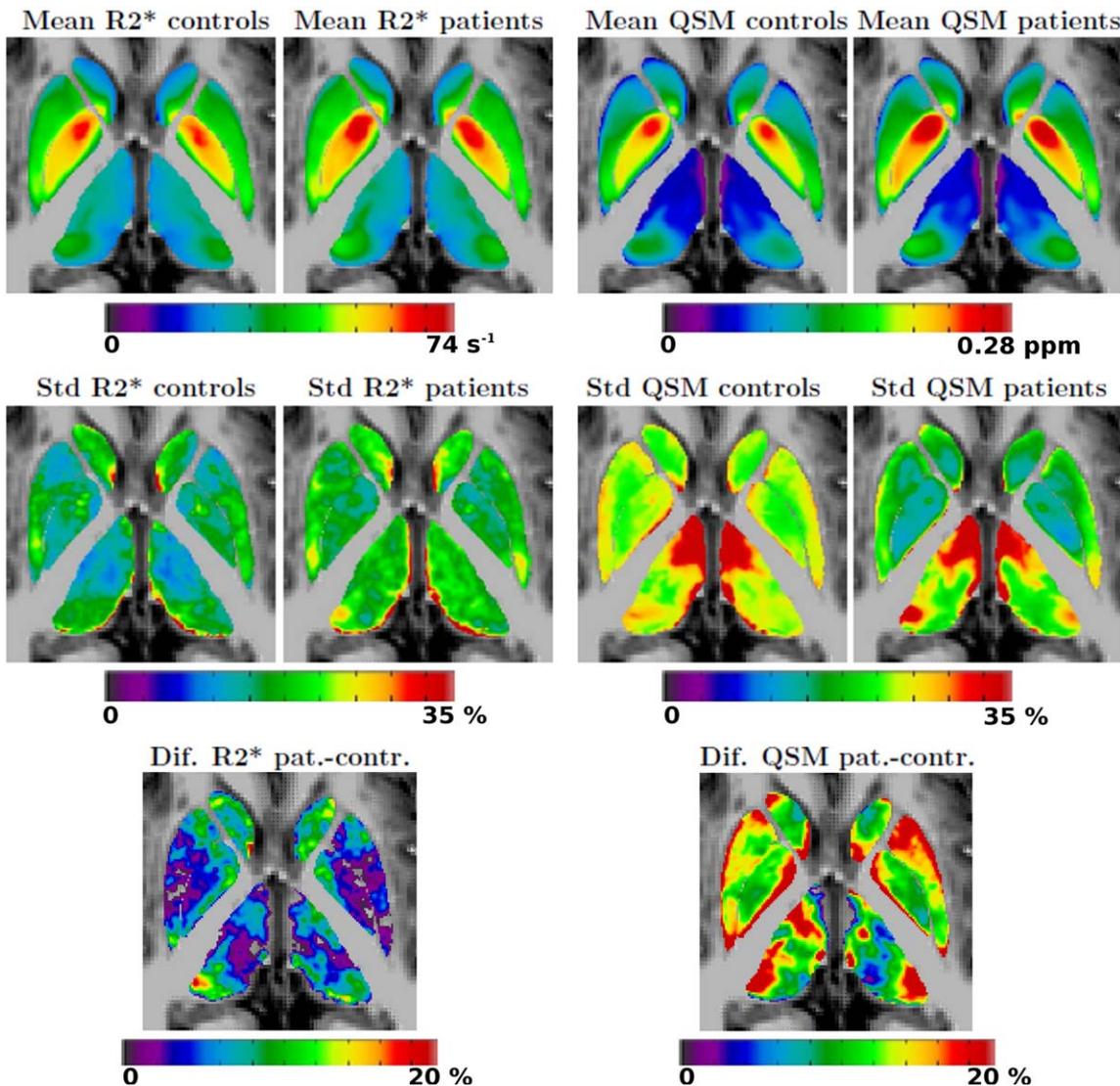

FIGURE 3: Comparison of group-averaged maps of R_2^* (left) and QSM (right) for MS patients and controls. Top row: Mean values averaged over controls and patients. Middle row: Mean normalized standard deviation in percent. Bottom row: Positive mean difference in percent of patients minus controls. Negative pixels within the deep GM structures are left blank without color.
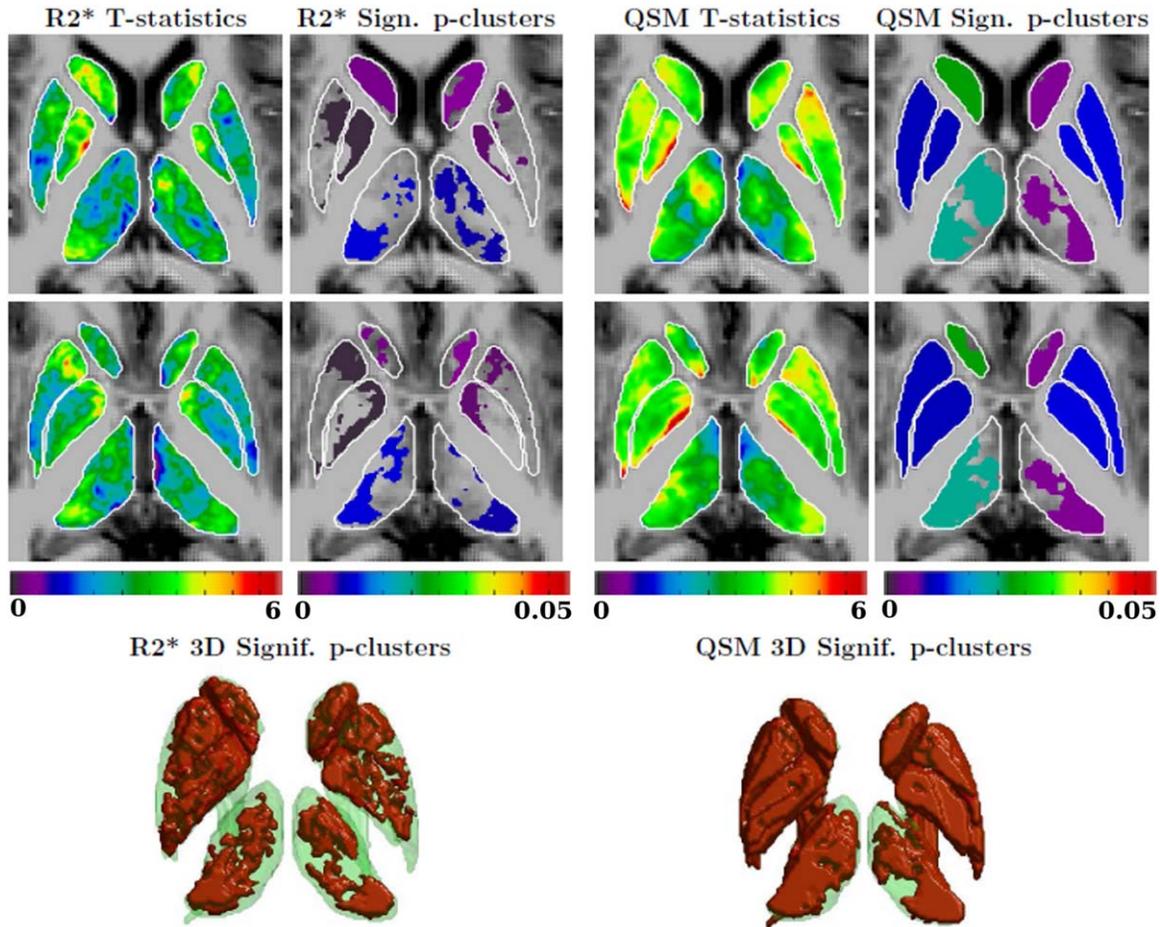
FIGURE 4: Statistical maps for group differences between patients and controls are shown of R_2^* (left) and QSM (right) for two slices, illustrating both *t*-statistics and significant random field corrected *P*-clusters ($P < 0.05$). A 3D view of all significant *P*-clusters is shown at the bottom.

Nevertheless, both sides have $P < 0.021$. The 3D views (bottom row) indicate all significant clusters across the volume.

Table 5 relates these 3D data to whole-structure measurements using volume-averaged entire bilateral structures compared to only the pixels of significance. Whole-structure measurements are found to have significant differences between patients and controls for R_2^* and QSM in all cases except caudate R_2^*. Whole-structure significance tends to be higher with QSM, as is evident in Table 5 and also Fig 4. However, when considering only significant pixels in Table 5, R_2^* produces higher levels of significance than QSM. Furthermore, the R_2^* significance increased at least 66-fold by limiting analysis to only significant pixels within each structure, while QSM only benefited 5-fold in the thalamus. Larger significant regions shown on QSM (Fig. 4) may be due to the regularization required for QSM, which creates internal smoothing (blurring) within structures.[3,24] Particularly for cluster analysis, this inherent smoothing of QSM may provide benefit, although mask focal variation. For example, from Table 5 almost the complete globus pallidus (99%) and putamen (97%) are significant for QSM, while only limited, focal territories are significant for R_2^* (46% and 35%, respectively).

Discussion

This study used QSM combined with standard T_1w images to achieve improved subcortical GM segmentation. The combination of QSM and T_1w images overcame the limited T_1w contrast of the globus pallidus and thalamus at 4.7T. With this improved segmentation, localized voxel-based analysis was performed in a control and MS population to examine regional variation in R_2^* and QSM. This localized analysis eliminates potential averaging and loss of significance from standard whole-structure analysis. Localized voxel-based analysis of R_2^* and QSM yielded significant differences between MS patients and controls in all four nuclei, with increased values in the patient group. These findings are consistent with previous work examining whole-structure R_2^* or QSM, which found increased R_2^* and QSM in MS subjects.[9,10,12–14]

QSM and R_2^* are complementary methods and both may be reconstructed when multiple gradient echo methods are used. Both measures are highly sensitive to iron, with R_2^* iron sensitivity increasing with field strength. For QSM, demyelination and iron accumulation both increase susceptibility, while for R_2^* they have opposing effects with

TABLE 5. Group-Averaged Mean R_2* and QSM From Whole Structures or Significant Regions[a]

Group/measure	Thalamus	Caudate	Putamen	Globus pallidus
Mean R_2* [s^{-1}]				
Patients	28.6 ± 2.8	29.3 ± 3.6	35.6 ± 4.6	50.3 ± 4.9
Controls	27.3 ± 2.0	28.3 ± 2.0	34.0 ± 3.1	47.5 ± 4.6
P-value	0.02*	0.07	0.03*	0.007*
Mean R_2* in significant regions [s^{-1}]				
% voxels	30%	42%	35%	46%
Patients	31.0 ± 3.4	31.6 ± 3.7	34.8 ± 4.4	50.3 ± 4.8
Controls	28.3 ± 2.5	29.0 ± 2.4	32.2 ± 2.5	46.1 ± 4.4
P-value	0.0003**	0.00003**	0.0003**	0.00004**
Mean QSM [ppm]				
Patients	0.065 ± 0.012	0.104 ± 0.013	0.107 ± 0.018	0.172 ± 0.020
Controls	0.058 ± 0.012	0.091 ± 0.018	0.092 ± 0.020	0.151 ± 0.032
P-value	0.01*	0.0005**	0.0008**	0.0008**
Mean QSM in significant regions [ppm]				
% voxels	57%	93%	97%	99%
Patients	0.073 ± 0.013	0.106 ± 0.014	0.108 ± 0.018	0.172 ± 0.020
Controls	0.063 ± 0.013	0.091 ± 0.018	0.093 ± 0.020	0.150 ± 0.032
P-value	0.002*	0.0003**	0.0007**	0.0008**

[a]All mean values reported as mean ± standard deviation.
*$P < 0.05$, **$P < 0.001$.

demyelination, or increased water content, reducing R_2*.[3–8] The increases in R_2* and QSM seen here are most likely due to increased iron accumulation in MS in the form of ferric iron stored in ferritin.[2–7] These increases in iron accumulation in MS versus healthy subjects need not be uniform within structures and may have focal regions with increased vulnerability, as evident particularly in the R_2* results. For example, within the thalamus the pulvinar region had a significant cluster on R_2*. Even in the cytologically uniform caudate and putamen, distinct clusters are seen on R_2*, but not on QSM. Further pathological and in vivo studies are necessary to gain a deeper understanding of within-structure variations in iron accumulation.

The most notable difference between R_2* and QSM was the size of the significant clusters, with QSM often encompassing the full structure, while R_2* results were more focal. Likely this is a result of the QSM reconstruction, which maintains boundaries but leads to blurring within the boundaries, creating a more uniform appearance within structures on QSM. This smoothing effect of QSM limits the value of localized voxel analysis over whole-structure analysis for QSM. Nevertheless, in the thalamus, voxel-based QSM eliminated regions of high standard deviation

such as the most anterior region, which has low QSM values. Moreover, the spatial blurring within QSM may be the main reason for the increased reliability over R_2*. For example, the inherent smoothing in QSM minimizes small vein effects that are bright on R_2*.

A previous study has examined R_2* and QSM voxel analysis in MS using whole-brain registration.[14] This study reported many areas of significance in white matter, but in deep GM significant voxels were only reported with QSM with no significant voxel-based results for R_2*, although whole-structure R_2* measures were significant. It is possible that R_2* is more sensitive to inaccurate registration than QSM, since it does not have inherent spatial regularization. Thus, the lack of voxel-based R_2* findings in their study[14] may have been limited by deep GM registration, since a whole-brain method was used and only T_1w images.

ROI measurements of deep GM in MS studies using R_2* or QSM have generally used either manual 2D ROIs[9,13,35] or automatic 3D ROIs using FSL FIRST.[10,12,14] Results with either method appear to be similar when the complete ROI is averaged. For voxel-based analysis, precise segmentation becomes more important than for large ROI analysis. Using whole 3D ROI analysis, we found significant

differences between MS patients and controls for all four nuclei with QSM, and all except caudate for R_2*. Restricting analysis to the significant voxels substantially improved significance for R_2*, but not for QSM. The most significant region of difference for R_2* was the globus pallidus, which also benefited most from improved segmentation over FSL FIRST. Previous work comparing QSM and R_2* using FSL FIRST at 3.0 or 7.0T[12,14] found no significance in the globus pallidus with R_2*, but only with QSM, possibly due to the difficulty of accurate globus pallidus segmentation, differences in patient groups, or different methods of determining R_2*.

There are numerous limitations to this work. A small number of subjects were studied, whereas a larger number of subjects could provide more statistical power and further elucidate specific abnormal areas. Some significant localized areas, especially along structural boundaries, may have not been identified due to segmentation, registration, and interpolation artifacts. Additionally, vascular contamination could increase the variance and compromise significance. The robustness of the segmentation system could further be improved by using more training data (manual segmentations). Within-boundary blurring effects from QSM reconstruction and the need for a relative background measure may have affected the results. Note that R_2* is an absolute measurement requiring no background normalization, while QSM requires a second background measure. For QSM, we used a white matter region as the background measure (internal capsule), due to variability in cerebrospinal fluid QSM measures. Lastly, longitudinal voxel-based analysis of QSM and R_2* may be required rather than the cross-sectional analysis employed here to further understand individual subject variation and to overcome the variable iron content found in healthy controls subjects.[31] Two year, whole ROI analysis of deep GM in MS using R_2* only has proven promising.[13]

In conclusion, automatic segmentation of iron-rich deep GM was improved by incorporating QSM along with T_1w images. Using this enhanced segmentation, detailed voxel-based analysis was possible, which yielded localized differences between RRMS subjects and controls within the globus pallidus, thalamus, caudate, and putamen using both R_2* and QSM. Localized analysis was of greater benefit to R_2* than to QSM. Further studies on localized iron accumulation in MS are warranted to examine variations within structures as well as longitudinal changes and treatment effects.

Acknowledgment

References

1. Schenck J, Zimmerman E. High-field MRI of brain iron: birth of a biomarker? NMR Biomed 2004;17:433–445.

2. Haacke E, Cheng N, House M, et al. Imaging iron stores in the brain using magnetic resonance imaging. Magn Reson Imaging 2005;23:1–25.

3. Wang Y, Liu T. Quantitative susceptibility mapping (QSM): decoding MRI data for a tissue magnetic biomarker. Magn Reson Med 2015 [Epub ahead of print].

4. Langkammer C, Krebs N, Goessler W, et al. Quantitative MR imaging of brain iron: a postmortem validation study. Radiology 2010;257:455–462.

5. Walsh AJ, Lebel RM, Eissa A, et al. Multiple sclerosis: validation of MR imaging for quantification and detection of iron. Radiology 2013;267:531–542.

6. Langkammer C, Schweser F, Krebs N, et al. Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study. Neuroimage 2012;62:1593–1599.

7. Zheng W, Nichol H, Liu S, et al. Measuring iron in the brain using quantitative susceptibility mapping and X-ray fluorescence imaging. Neuroimage 2013;78:68–74.

8. Sun H, Walsh AJ, Lebel RM, et al. Validation of quantitative susceptibility mapping with Perls' iron staining for subcortical gray matter. Neuroimage 2015;105:486–492.

9. Lebel RM, Eissa A, Blevins G, Wilman AH. Quantitative magnetic resonance imaging of subcortical gray matter in multiple sclerosis. Mult Scler 2012;18:433–441.

10. Khalil M, Langkammer C, Ropele S, et al. Determinants of brain iron in multiple sclerosis: a quantitative 3T MRI study. Neurology 2011;77:1691–1697.

11. Zivadinov R, Heininen-Brown M, Schirda CV, et al. Abnormal subcortical deep-gray matter susceptibility-weighted imaging filtered phase measurements in patients with multiple sclerosis: a case-control study. Neuroimage 2012;59:331–339.

12. Langkammer C, Liu T, Khalil M, et al. Quantitative susceptibility mapping in multiple sclerosis. Radiology 2013;267:551–559.

13. Walsh AJ, Blevins G, Lebel RM, et al. Longitudinal MR imaging of iron in multiple sclerosis: an imaging marker of disease. Radiology 2014;270:186–196.

14. Rudko DA, Solovey I, Gati JS, Kremenchutzky M, Menon RS. Multiple sclerosis: improved identification of disease-relevant changes in gray and white matter by using susceptibility-based MR imaging. Radiology 2014;272:851–864.

15. Derakhshan M, Caramanos Z, Giacomini PS, et al. Evaluation of automated techniques for the quantification of gray matter atrophy in patients with multiple sclerosis. NeuroImage 2010;52:1261–1267.

16. Gelman N, Gorell J, Barker P, et al. Preliminary report on transverse relaxation rates and relation to estimated iron content. Radiology 1999;210:759–767.

17. Yao, B, Li, TQ, Gelderen, P, et al. Susceptibility contrast in high field MRI of human brain as a function of tissue iron content. NeuroImage 2009;44:1259–1266.

18. Lim IA, Faria AV, Li X, Hsu JT, Airan RD, Mori S, van Zijl PC. Human brain atlas for automated region of interest selection in quantitative susceptibility mapping: application to determine iron content in deep gray matter structures. Neuroimage 2013;82:449–469.

19. Heckemann R, Hajnal J, Aljabar P, et al. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 2006;33:115–126.

20. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage 2012;62:782–790.

21. Sun H, Wilman AH. Background field removal using spherical mean value filtering and Tikhonov regularization. Magn Reson Med 2014;71:1151–1157.

22. Bilgic B, Pfefferbaumb A, Rohlfing T, et al. MRI estimates of brain iron concentration in normal aging using quantitative susceptibility mapping. Neuroimage 2012;59:2625–2635.

23. Liu T, Liu J, de Rochefort L, et al. Morphology enabled dipole inversion (MEDI) from a single-angle acquisition: comparison with COSMOS in human brain imaging. Magn Reson Med 2011;66:777–783.

24. Wu B, Li W, Guidon A, Liu C. Whole brain susceptibility mapping using compressed sensing. Magn Reson Med 2012;67:137–147.

25. FreeSurfer: http://surfer.nmr.mgh.harvard.edu/ (2011).

26. Polman C, Reingold S, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the Mcdonald criteria. Ann Neurol 2005;58:840–846.

27. Babalola KO, Patenaude B, Aljabar P, et al. An evaluation of four automatic methods of segmenting the subcortical structures in the brain. NeuroImage 2009;47:1435–1447.

28. Rasband W. Imagej: http://rsbweb.nih.gov/ij/ (1997–2014).

29. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 2011;54:2033–2044.

30. Warfield S, Zou K, Wells W. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004;23:903–921.

31. Hallgren B, Sourander P. The effect of age on the non-haemin iron in the human brain. J Neurochem 1958;3:41–51.

32. Cherubini A, Peran P, Caltagirone C, et al. Aging of subcortical nuclei: microstructural, mineralization and atrophy modifications measured in vivo using MRI. NeuroImage 2009;48:29–36.

33. Worsley KJ. Surfstat: http://www.math.mcgill.ca/keith/surfstat (2008).

34. Adler RJ, Taylor JE. Random fields and geometry. Berlin: Springer; 2007.

35. Al-Radaideh AM, Wharton SJ, Lim SY, et al. Increased iron accumulation occurs in the earliest stages of demyelinating disease: an ultra-high field susceptibility mapping study in clinically isolated syndrome. Mult Scler 2013;19:896–903.