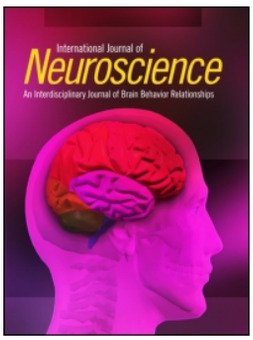

International Journal of Neuroscience

ISSN: 0020-7454 (Print) 1543-5245 (Online) Journal homepage: http://www.tandfonline.com/loi/ines20

Sample size requirements for one-year treatment effects using deep gray matter volume from 3T MRI in progressive forms of multiple sclerosis

Gloria Kim, Renxin Chu, Fawad Yousuf, Shahamat Tauhid, Lynn Stazzone, Maria K. Houtchens, James M. Stankiewicz, Christopher Severson, Dorlan Kimbrough, Francisco J. Quintana, Tanuja Chitnis, Howard L. Weiner, Brian C. Healy & Rohit Bakshi

To cite this article: Gloria Kim, Renxin Chu, Fawad Yousuf, Shahamat Tauhid, Lynn Stazzone, Maria K. Houtchens, James M. Stankiewicz, Christopher Severson, Dorlan Kimbrough, Francisco J. Quintana, Tanuja Chitnis, Howard L. Weiner, Brian C. Healy & Rohit Bakshi (2017): Sample size requirements for one-year treatment effects using deep gray matter volume from 3T MRI in progressive forms of multiple sclerosis, International Journal of Neuroscience, DOI: 10.1080/00207454.2017.1283313

To link to this article: http://dx.doi.org/10.1080/00207454.2017.1283313

Accepted author version posted online: 19 Jan 2017.
Published online: 02 Feb 2017.

INTERNATIONAL JOURNAL OF NEUROSCIENCE, 2017
http://dx.doi.org/10.1080/00207454.2017.1283313

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS

Sample size requirements for one-year treatment effects using deep gray matter volume from 3T MRI in progressive forms of multiple sclerosis

Gloria Kim[a], Renxin Chu[a], Fawad Yousuf[a], Shahamat Tauhid[a], Lynn Stazzone[a], Maria K. Houtchens[a], James M. Stankiewicz[a], Christopher Severson[a], Dorlan Kimbrough[a], Francisco J. Quintana[a], Tanuja Chitnis[a], Howard L. Weiner[a], Brian C. Healy[a] and Rohit Bakshi [a,b]

[a]Departments of Neurology Brigham and Women's Hospital, Laboratory for Neuroimaging Research, Partners MS Center, Harvard Medical School, Boston, MA, USA; [b]Radiology Brigham and Women's Hospital, Laboratory for Neuroimaging Research, Partners MS Center, Harvard Medical School, Boston, MA, USA

ABSTRACT

Objective: The subcortical deep gray matter (DGM) develops selective, progressive, and clinically relevant atrophy in progressive forms of multiple sclerosis (PMS). This patient population is the target of active neurotherapeutic development, requiring the availability of outcome measures. We tested a fully automated MRI analysis pipeline to assess DGM atrophy in PMS.

Design/Methods: Consistent 3D T1-weighted high-resolution 3T brain MRI was obtained over one year in 19 consecutive patients with PMS [15 secondary progressive, 4 primary progressive, 53% women, age (mean±SD) 50.8±8.0 years, Expanded Disability Status Scale (median, range) 5.0, 2.0–6.5)]. DGM segmentation applied the fully automated FSL-FIRST pipeline (http://fsl.fmrib.ox.ac.uk). Total DGM volume was the sum of the caudate, putamen, globus pallidus, and thalamus. On-study change was calculated using a random-effects linear regression model.

Results: We detected one-year decreases in raw [mean (95% confidence interval): –0.749 ml (–1.455, –0.043), $p = 0.039$] and annualized [–0.754 ml/year (–1.492, –0.016), $p = 0.046$] total DGM volumes. A treatment trial for an intervention that would show a 50% reduction in DGM brain atrophy would require a sample size of 123 patients for a single-arm study (one-year run-in followed by one-year on-treatment). For a two-arm placebo-controlled one-year study, 242 patients would be required per arm. The use of DGM fraction required more patients. The thalamus, putamen, and globus pallidus, showed smaller effect sizes in their on-study changes than the total DGM; however, for the caudate, the effect sizes were somewhat larger.

Conclusions: DGM atrophy may prove efficient as a short-term outcome for proof-of-concept neurotherapeutic trials in PMS.

ARTICLE HISTORY
Received 16 December 2016
Revised 6 January 2017
Accepted 13 January 2017
Published online 2 February 2017

KEYWORDS
Multiple sclerosis; deep gray matter; MRI; 3T; progressive MS; brain atrophy

Introduction

MRI provides a powerful tool for the diagnosis and longitudinal monitoring of multiple sclerosis (MS) [1]. Standard supportive MRI outcome measures employed in clinical therapeutic trials [2] include quantitative assessments of cerebral lesions [3] and atrophy [4]. However, in patients with advanced disability and progressive forms of the disease, lesion measures, both the accumulated burden and new activity, are less sensitive to disease progression, than in patients with early stage relapsing forms of the disease [5]. Given that both primary progressive (PP) and secondary progressive (SP) MS represent a major unmet therapeutic opportunity [6] and an active area of therapeutic development [7,8], there is an urgent need to develop efficient biomarkers to track such responses.

Brain volume is a leading proposed tool to assess neuroprotection in MS trials [9], including those in progressive forms of MS [10], owing in part to its close relationship to several clinical aspects of MS [4] including physical disability [11], cognitive impairment [12], quality of life [13], and employment status [14]. The cerebral gray matter (GM) develops atrophy throughout the MS disease course, both very early [15] and as the disease continues to advanced stages [16]. GM and white matter (WM) pathology are not necessarily closely coupled [15]. Furthermore, previous studies have shown that the deep gray matter (DGM, i.e., basal ganglia and thalamus) is disproportionately affected as compared to the cortical GM [17] and whole brain [18,19] and such involvement is highly clinically relevant [19–22]. Regarding the methodology available to assess brain volume loss from MRI

CONTACT Rohit Bakshi rbakshi@post.harvard.edu

Table 1. Patient demographic and clinical data.

	Multiple sclerosis patients ($n = 19$)
Age (years) (mean ± SD) (range)	50.8 ± 8.0 (22.9, 57.4)
Women, *n* (%)	10 (53%)
Disease category, *n* (%)	
Secondary progressive	15 (79%)
Primary progressive	4 (21%)
Disease duration[a] (years) (mean ± SD) (range)	17.9 ± 10.0 (5.6, 39.6)
Clinical data	
–Baseline	
EDSS score (median, range)	5.0 (2.0, 6.5)
T25FW (median, range)	6.3 (3.8, 68.0)
–One-year follow-up	
EDSS score (median, range)	6.0 (2.0, 7.5)
T25FW (median, range)	6.7 (4.0, 55.5)[b]

Note: SD = standard deviation.
[a]Since first symptoms; EDSS = Expanded Disability Status Scale; T25FW = timed 25-foot walk, seconds.
[b]One subject could not walk at follow-up and therefore T25FW data are $n = 18$.

scans, recent data indicate advantages of 3T vs. 1.5T MRI for accuracy and validity [23], which parallels its advantages in assessing lesions [24–26].

Against this backdrop, we applied high-resolution 3T images to a fully automated analysis pipeline to assess the ability to detect one year of DGM atrophy in patients with progressive forms of MS (PMS). Furthermore, to show the potential application of these results to therapeutic monitoring, we also calculated sample size requirements for two study designs regarding the demonstration of neuroprotective treatment effects.

Subjects and methods

Subjects

All subjects' baseline demographic and clinical data are presented in Table 1. This was a retrospective study of 19 consecutive patients from the Partners MS Center. Patients were required to have PMS, either PP or SP [27], with an age of 18–60 years. According to the latest MS diagnostic guidelines [27], the patients were further characterized as: active and with progression = 1; active but without progression = 0; not active but with progression = 7; not active and without progression (stable disease) = 11. Patients were also required to have undergone brain MRI at baseline and one year later using a consistent acquisition protocol on the same scanner both within and across patients. One-year follow-up MRI was 8-14 (mean 11.5) months from baseline. None of the patients received corticosteroids in the four weeks prior to MRI. Each patient was assessed at both MRI time points by an MS specialist neurologist using the Expanded Disability Status Scale (EDSS) [28] and timed 25-foot walk (T25FW) [29]. During the one-year study, as prescribed in a "real world" clinical setting by the neurologist, 14 patients (74%) were receiving monotherapy with disease-modifying treatment (DMT): rituximab ($n = 5$), fingolimod ($n = 1$), dimethyl fumarate ($n = 2$), natalizumab ($n = 3$), teriflunomide ($n = 1$), glatiramer acetate ($n = 1$), and mycophenolate mofetil ($n = 1$). One patient was receiving both glatiramer acetate and rituximab in combination. Another patient was receiving dimethyl fumarate at the beginning of the study for 2.5 months, then no DMT for 3.5 months, followed by natalizumab for the remainder of the study period. Thus, 84% (16/19) of patients received DMT during the study period while 16% ($n = 3$) did not. This work was part of the Comprehensive Longitudinal Investigation of Multiple Sclerosis at the Brigham and Women's Hospital (CLIMB) study [30], requiring informed consent, and was approved by The Partners Human Research Committee, our institutional review board.

MRI acquisition and analysis

For all patients, the brain MRI acquisition platform and protocol applied to obtain high-resolution 3D T1-weighted images with isotropic voxels is shown in Table 2. Image pre-processing was performed using Jim 7.0 (Xinapse Systems Ltd., Northants, UK, http://www.xinapse.com/). First, all original DICOM images were converted to a Neuroimaging Informatics Technology Initiative (NIfTI) format, and their raw sagittal orientation was converted to axial. Then, 170 axial slices were extracted from each scan starting at the first slice showing the top of the head. This provided whole brain coverage in all patients extending to the foramen magnum. All other additional slices were removed. The images were then applied to a fully automated pipeline (FSL-FIRST, v. 5.0, The FMRIB Analysis Group, Oxford, UK, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIRST) to derive volumes of the caudate, putamen, globus pallidus, and thalamus (Figure 1). Lesion filling was not performed to maintain full automation of the pipeline. Total DGM volume was the sum of

Table 2. Brain 3T MRI acquisition protocol.

Scanner manufacturer and model	Siemens Skyra
Type of sequence	3D MPRAGE
Orientation	Sagittal
Field of view (cm)	24 × 25.6
Matrix size	240 × 256
Number of slices	176
Repetition time (msec)	2300
Echo time (msec)	2.96
Inversion time (msec)	900
Flip angle (degrees)	9
Voxel size (mm)	1.0 × 1.0 × 1.0
Scan time (minutes)	5.2
Number of signal averages	1
Coil	20 channel craniospinal
Gradient strength	45 mT/m @ 200 T/m/s

Note: MPRAGE = magnetization-prepared rapid acquisition gradient echo.

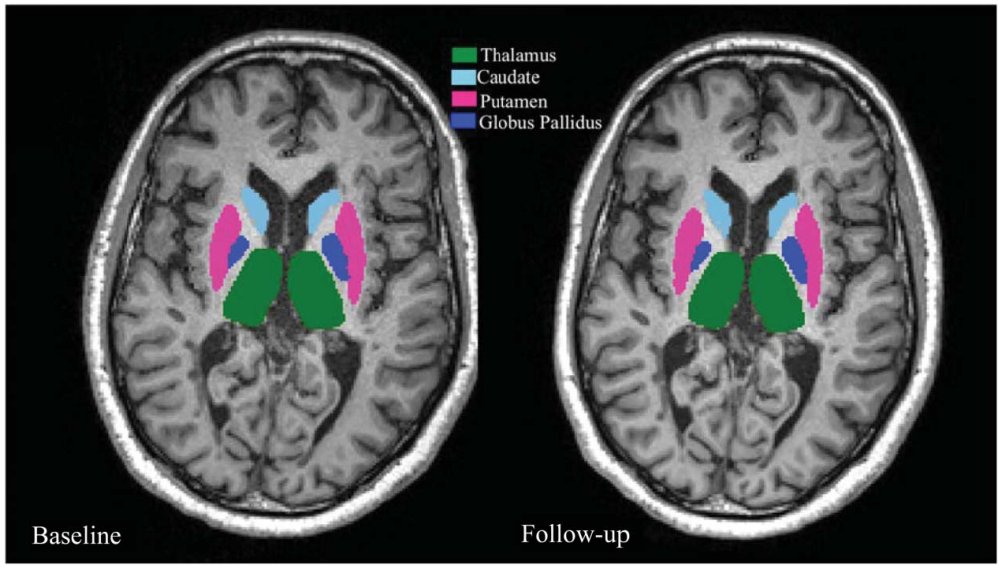

Figure 1. Examples of deep gray matter segmentation from MRI scans.
Raw 3D T1-weighted images that have been re-sampled to axial with examples of FSL/FIRST (Oxford Centre for Functional MRI of the Brain Software Library Integrated Registration and Segmentation Tool) deep gray matter segmentation. In the present study, we utilized the FSL/FIRST outputs to assess the volume of the thalamus, caudate, putamen, and globus pallidus (and their sum = total deep gray matter). Images are from one patient at baseline (left) and one-year follow-up (right).

these structures. In addition to raw volumes, we also calculated fractions, by dividing each volume by the intracranial volume (ICV), an estimate of the skull cavity commonly used to normalize brain volume measures [31]. ICV was obtained by applying the 3D T1-weighted images to a fully automated segmentation-based algorithm, (SIENAX, v. 5.0, The Analysis Group, Oxford, UK, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/SIENA), based on our previous work in optimizing this tool for 3T images [23]. In addition, the patients had clinically relevant uniform brain MRI acquisitions to assess gadolinium-enhanced lesions at both time points (T1-weighted axial images, pre- and post-contrast) and new/enlarging T2 lesions at follow-up (fluid-attenuated inversion-recovery and T2-weighted axial images).

Reliability analysis

As we reported recently in a meeting presentation [32], the DGM volume and fraction measures from this platform and acquisition sequence were highly reliable in patients with MS ($n = 14$), based on scan-rescan experiments, with intraclass correlation coefficients (ICCs) ranging from 0.92 to 1.00. Raw (non-normalized) volumes showed higher reliability (ICCs 0.98–1.00) than fractions (ICCs 0.92–0.95). Furthermore, total DGM volume was measured with higher reliability (ICC 0.99–1.00) than the individual DGM structure volumes (ICCs 0.97–0.99). These reliability data have been submitted in full detail as part of a separate manuscript.

Statistical assumptions/analysis

The current sample represents a "real world" comparison treatment arm that, while heterogeneous, we propose could be considered similar to a "placebo"; this is in light of the observations that the DMTs utilized in this study have not shown a therapeutic benefit for the primary clinical outcome measure or brain atrophy in phase III clinical trials of patients with progressive forms of MS [6, 33, 34]. Thus, even if a hypothetical trial considered with our power calculations had an active comparison arm with one of the currently available DMTs, it could be considered a de facto placebo for the calculations. Overall, we feel this is a conservative approach in that if we assume that the current treatments might slow GM atrophy, we can view our sample size calculations as potentially an overestimate of the true required size.

The on-study change in each measure was calculated using a random-effects linear regression model. The raw on-study change was calculated using a categorical time variable, and the annualized change was calculated using continuous time for the time between the two MRI scans. Given the small sample size and the non-uniform follow-up intervals, we suggest the annualized results are more coherent, would be better applied to a clinical trial setting, and should be given more emphasis in the interpretation of our data. Using the estimated means and residual standard deviation from the mixed model, sample size calculations for two types of trials were completed. First, we calculated the sample size for a placebo-controlled trial in which the goal was to show that there

was a significant difference in the change from baseline between the two groups. For this calculation, the change in the placebo group was assumed to equal the mean change observed in our sample, and we assumed a treatment effect that reduced the change by 50%. Second, we calculated the sample size for a one-arm trial in which the change over one year prior to treatment was compared to the change over one year on treatment. For this calculation, the change in the pre-treatment period was assumed to equal the mean change observed in our sample; we assumed a treatment effect that reduced the change by 50%; and we assumed a correlation between the change pre-treatment and the change on treatment of 0.5. We note that higher correlations between the pre-treatment and on treatment change measurements would lead to a smaller sample size and vice versa. We also calculated sample sizes for a 25% treatment effect.

We note that for both calculations, the *power two means* and *power paired means*, routines in Stata v14 (Stata Corp LP. College Station, Texas, http://www.stata.com/) were used, and the variance for these calculations was equal to two times the residual variance from the random-effects model. In all sample size calculations, we used a power of 80%. Finally, the on-study change in EDSS was assessed using a random-effects proportional odds logistic regression model.

Results

On-study MRI changes – total DGM

On-study MRI changes in DGM are shown in Tables 3 and 4, Figures 2–7. Regarding total DGM, significant decreases in volumes [mean (95% CI): –0.749 ml (–1.455, –0.043), $p = 0.039$] and annualized [–0.754 ml/year (–1.492, –0.016), $p = 0.046$] volumes were detected. Decreases in raw [–0.000359 (–0.000800, 0.000083), (–1.37%) $p = 0.105$] and annualized [–0.00345 (–0.000807, 0.000117), $p = 0.134$] total DGM fractions were detected over one year, which approached but did not reach statistical significance. The total DGM volume decreased by an average of 2.2% across all patients from baseline to follow-up. We also tested two additional methods of normalization of DGM volumes: the residual method using ICV [31] and multiplying each raw volume by the whole brain scaling factor from SIENAX. For annualized change using either of these methods, there were generally smaller effect sizes than the raw data for the DGM volumes (data not shown).

Table 3. Baseline and on-study changes in cerebral deep gray matter raw volumes and fractions: actual (non-annualized) data and sample size calculations for a 50% treatment effect.

MRI measure	Baseline: mean (SD)	Follow-up: mean (SD)	% change: mean (SD)	Estimated change (95% CI)	Residual standard deviation (95% CI)	*p* Value	Sample size – two-arm study[A]	Sample size – one-arm study[B]
Raw thalamus	14.73 (1.68)	14.42 (1.68)	−1.99 (4.65)	−0.307 (−0.647, 0.033)	0.499 (0.360, 0.692)	0.074	668	168
Raw caudate	6.21 (0.98)	6.11 (0.94)	−1.57 (2.56)	−0.103 (−0.188, −0.019)	0.124 (0.089, 0.172)	0.020*	366	93
Raw putamen	9.13 (1.09)	8.90 (1.24)	−2.43 (8.05)	−0.233 (−0.586, 0.121)	0.519 (0.374, 0.719)	0.184	1250	314
Raw globus pallidus	3.27 (0.37)	3.16 (0.42)	−3.18 (7.47)	−0.106 (−0.222, 0.011)	0.171 (0.124, 0.238)	0.074	658	166
Raw DGM	33.34 (3.85)	32.59 (3.92)	−2.19 (4.43)	−0.749 (−1.455, −0.043)	1.036 (0.747, 1.436)	0.039*	484	123
Fraction thalamus	0.0116 (0.0008)	0.0115 (0.0009)	−1.12 (3.34)	−0.000130 (−0.000327, 0.000066)	0.000288 (0.000208, 0.000399)	0.180	1234	310
Fraction caudate	0.0049 (0.0006)	0.0049 (0.0005)	−0.61 (3.51)	−0.000034 (−0.00012, 0.000056)	0.000132 (0.000095 0.000182)	0.437	3784	948
Fraction putamen	0.0072 (0.0006)	0.0071 (0.0007)	−1.56 (7.53)	−0.000132 (−0.000410, 0.000146)	0.000408 (0.000294,0.000565)	0.333	2394	600
Fraction globus pallidus	0.0026 (0.0002)	0.0025 (0.0002)	−2.37 (6.17)	−0.0000629 (−0.000140, 0.000014)	0.000113 (0.000082, 0.000157)	0.103	816	206
Fraction DGM	0.0263 (0.0019)	0.0260 (0.0019)	−1.30 (3.33)	−0.000359 (−0.000800, 0.000083)	0.000648 (0.000467, 0.000898)	0.105	820	207

Note: *$p < 0.05$. Sample sizes represent the total number of patients required to test a therapy showing a 50% reduction in the rate of atrophy. A = one-year placebo-controlled parallel group (two groups, treatment vs. placebo). B = single-arm study, with a one-year pre-treatment run-in followed by a one-year treatment period. Variance for each sample size calculation was residual variance times 2. DGM = Deep gray matter (thalamus+caudate+putamen+globus pallidus). Raw volumes are in ml.

Table 4. On-study changes in cerebral deep gray matter raw volumes and fractions: annualized data: sample size calculations for a 50% treatment effect.

Measure	Estimated change (95% CI)	Residual standard deviation (95% CI)	*p* Value	Sample size – two-arm study[A]	Sample size – one-arm study[B]
Raw thalamus	−0.319 (−0.672, 0.034)	0.501 (0.361, 0.694)	0.074	622	157
Raw caudate	−0.101 (−0.190, −0.011)	0.127 (0.092, 0.176)	0.030*	400	101
Raw putamen	−0.223 (−0.591, 0.146)	0.523 (0.377, 0.725)	0.221	1382	347
Raw globus pallidus	−0.111 (−0.232, 0.0099)	0.172 (0.124, 0.238)	0.070	604	153
Raw deep gray matter	−0.754 (−1.492, −0.016)	1.047 (0.755, 1.451)	0.046*	486	123
Fraction thalamus	−0.000131 (−0.000335, 0.000072)	0.000289 (0.000208, 0.000401)	0.192	1224	308
Fraction caudate	−0.000031 (−0.000124, 0.000062)	0.000132 (0.000095, 0.000183)	0.495	4572	1145
Fraction putamen	−0.000108 (−0.000397, 0.000182)	0.00041 (0.000296, 0.000569)	0.445	3624	908
Fraction globus pallidus	−0.0000648 (−0.000145, 0.000015)	0.000113 (0.000082, 0.000157)	0.105	768	194
Fraction deep gray matter	−0.000345 (−0.000807, 0.000117)	0.000655 (0.000472, 0.000907)	0.134	908	229

Note: *$p < 0.05$. Sample sizes represent the total number of patients required to test a therapy showing a 50% reduction in the rate of atrophy. A = one-year placebo-controlled parallel group (two groups, treatment vs. placebo). B = single-arm study, with a one-year pre-treatment run-in followed by a one-year treatment period. Variance for each sample size calculation was residual variance times 2. Deep gray matter = thalamus+caudate+putamen+globus pallidus. Raw volume changes are in ml.

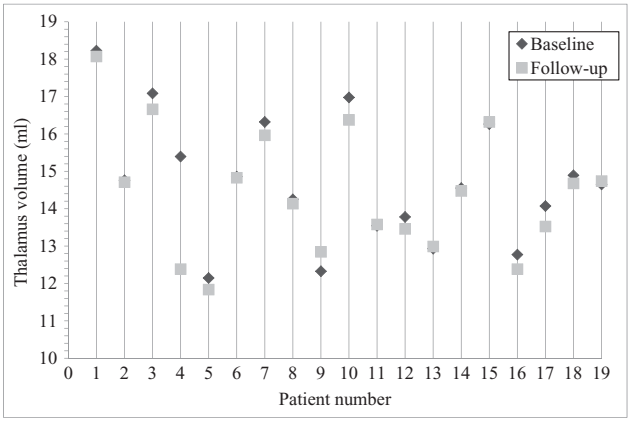

Figure 2. Thalamus volume change in each subject.
Individual patient data showing baseline (diamonds) and one-year follow-up (squares) raw thalamic volume at baseline and follow-up.

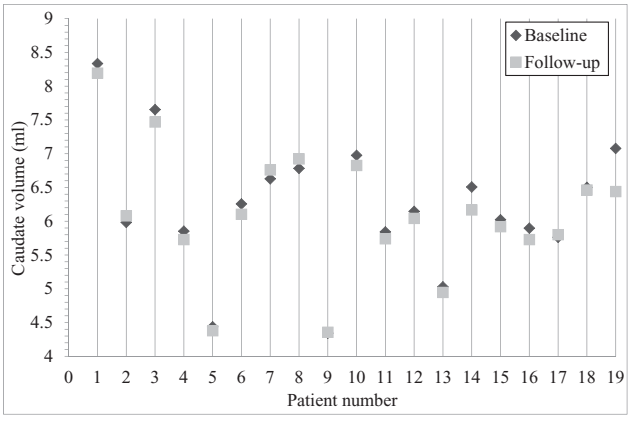

Figure 3. Caudate volume change in each subject.
Individual patient data showing baseline (diamonds) and one-year follow-up (squares) caudate volume at baseline and follow-up.

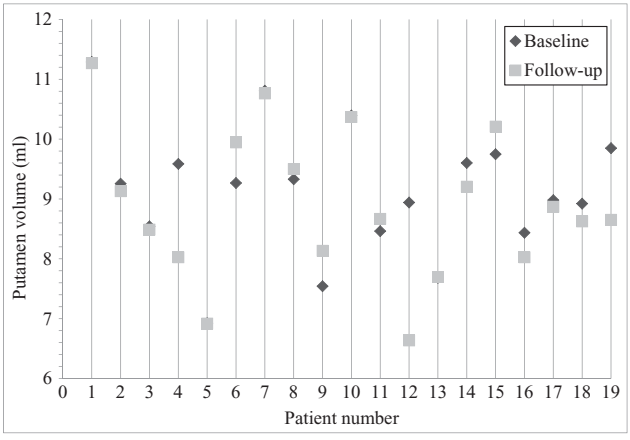

Figure 4. Putamen volume change in each subject.
Individual patient data showing baseline (diamonds) and one-year follow-up (squares) putamen volume at baseline and follow-up.

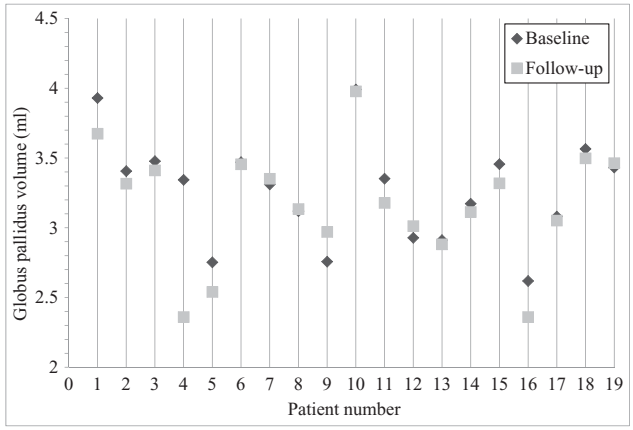

Figure 5. Globus pallidus change in each subject.
Individual patient data showing baseline (diamonds) and one-year follow-up (squares) globus pallidus volume at baseline and follow-up.

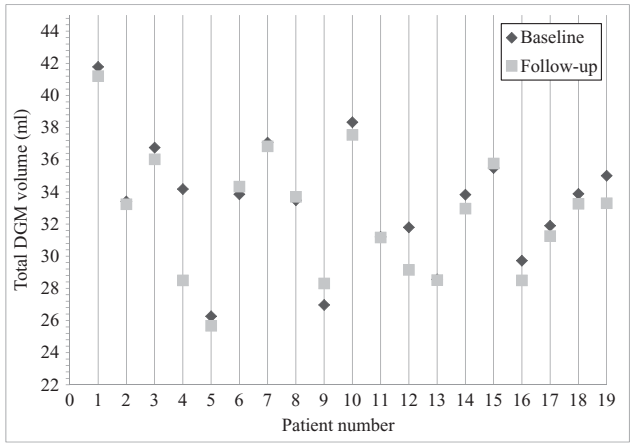

Figure 6. Total cerebral deep gray matter change in each subject.
Individual patient data showing baseline (diamonds) and one-year follow-up (squares) total cerebral deep gray matter (DGM = thalamus+caudate+putamen+globus pallidus) volume at baseline and follow-up.

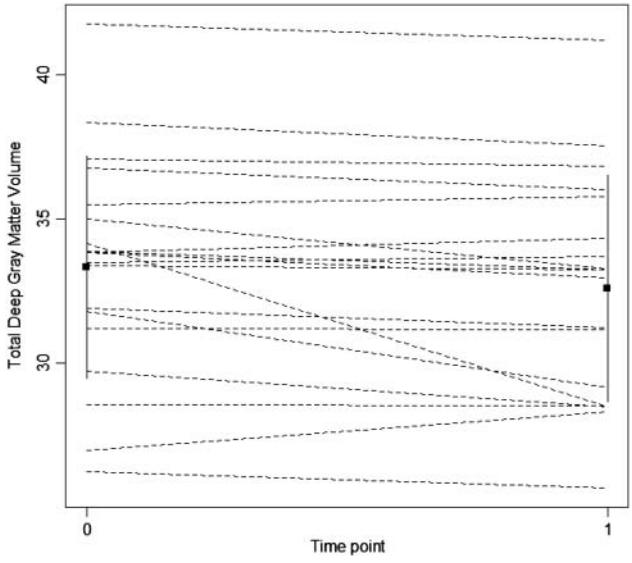

Figure 7. Total deep gray matter one-year change in all subjects.
Total cerebral deep gray matter volume (ml) on-study in the cohort, showing means (solid squares), standard deviation bars, and individual patients' data (dotted lines). Time points are presented in years.

Table 5. On-study changes in cerebral deep gray matter raw volumes: actual (non-annualized) data: sample size calculations for a 25% treatment effect.

MRI measure	Estimated change (95% CI)	Residual standard deviation (95% CI)	*p* Value	Sample size – two-arm study[A]	Sample size – one-arm study[B]
Raw thalamus	−0.307 (−0.647, 0.033)	0.499 (0.360, 0.692)	0.074	2660	667
Raw caudate	−0.103 (−0.188, −0.019)	0.124 (0.089, 0.172)	0.020*	1456	365
Raw putamen	−0.233 (−0.586, 0.121)	0.519 (0.374, 0.719)	0.184	4988	1249
Raw globus pallidus	−0.106 (−0.222, 0.011)	0.171 (0.124, 0.238)	0.074	2622	657
Raw deep gray matter	−0.749 (−1.455, −0.043)	1.036 (0.747, 1.436)	0.039*	1924	483

Note: *$p < 0.05$. Sample sizes represent the total number of patients required to test a therapy showing a 25% reduction in the rate of atrophy. A = one-year placebo-controlled parallel group (two groups, treatment vs. placebo). B = single-arm study, with a one-year pre-treatment run-in followed by a one-year treatment period. Variance for each sample size calculation was residual variance times 2. Deep gray matter = thalamus+caudate+putamen+globus pallidus. Raw volume changes are in ml.

Sample size requirements using total DGM as an outcome measure are shown in Tables 3 and 4. Considering a hypothetical treatment trial for an intervention that would show a 50% reduction in the loss of DGM, the sample size requirement for a single-arm study was 123 patients using total DGM volume. For a two-arm placebo-controlled parallel-group one-year study, more than 200 patients would be required per arm. When using total DGM fraction, the sample size requirements were higher for both types of clinical trials. We also present data for hypothetical trial showing a 25% treatment effect - the sample size requirement for a single-arm study was 483 patients using total DGM volume; for a two-arm study, more than 950 patients per arm would be required (Table 5).

On-study MRI changes – individual DGM nuclei

An examination of the changes in raw volumes and fractions of the thalamus, putamen, and globus pallidus, showed decreases over time but with smaller effect sizes in their on-study changes than the total DGM data; however, for the caudate, the effect sizes for the decreases over time were somewhat larger (Tables 3 and 4; Figures 2–5). The volume decreased by an average of 2.1% for thalamus, 1.7% for the caudate, 2.5% for putamen, and 3.2% for globus pallidus across all patients from baseline to follow-up. Globus pallidus volume changes had a smaller effect size despite the fact that its percent volume decrease was the highest. In contrast, the effect size was larger with the caudate, while its percent volume decrease was the lowest. This is because the mixed model results and calculated sample size requirements are based on both the estimated change and the standard deviation of the change. Table 3 shows that the change in the caudate was small, but the standard deviation was also quite small. There was one subject with a marked decrease in all measures during the observation period (subject 4) compared to the other subjects (Figures 2–6). When this subject was removed from the analyses, the estimated change decreased for the outcomes, but the standard error and residual standard deviation decreased. Therefore, the *p*-values for the estimated changes and the estimated sample sizes based on the new values were similar between this subset and the full sample (data not shown).

Baseline and on-study MRI lesion activity

At baseline, one patient had an active MRI scan with a single gadolinium-enhancing; none of the other patients had active lesions at baseline. At follow-up, none of the patients had new/enlarging T2 lesions or enhancing lesions.

Exploratory analysis: subgroup differences

Although sample size was limited, we explored differences in DGM volume changes based on patient characteristics (Tables 6 and 7). Patients with SP MS ($n = 15$) had a lower rate of atrophy in the thalamus and globus pallidus as compared to the patients with PP MS ($n = 4$). There were no differences in on-study DGM volume change between patients receiving immunotherapy ($n = 16$) and those untreated ($n = 3$).

On-study clinical changes

Baseline and one-year follow-up clinical data are summarized in Table 1. The EDSS scores did not change between baseline and follow-up ($p = 0.19$).

Table 6. On-study changes in cerebral deep gray matter raw volumes and fractions: primary ($n = 4$) vs. secondary ($n = 15$) progressive patients with MS.

Measure	Estimated difference in change in raw volume (95% CI)	*p* Value	Estimated difference in change in fractional volume (95% CI)	*p* Value
Thalamus	1.05 (0.224, 1.873)	0.016*	0.00051 (0.00001, 0.00101)	0.045*
Caudate	−0.006 (−0.254, 0.242)	0.960	−0.00016 (−0.00040, 0.00009)	0.189
Putamen	0.438 (−0.559, 1.435)	0.367	0.00011 (−0.00069, 0.00091)	0.771
Globus pallidus	0.407 (0.142, 0.672)	0.005*	0.00025 (0.00007, 0.00043)	0.009*
Deep gray matter	1.88 (0.072, 3.697)	0.042*	0.00075 (−0.00048, 0.0020)	0.215

Note: The estimated coefficient is the difference in the annualized change over time in the secondary progressive vs. primary progressive group. A positive number implies less decrease in volume in the secondary progressive subjects. Deep gray matter = thalamus+caudate+putamen+globus pallidus. Raw volume changes are in ml

*$p < 0.05$.

Table 7. On-study changes in cerebral deep gray matter raw volumes and fractions: treated ($n = 16$) vs. untreated ($n = 3$) progressive patients with MS.

Measure	Estimated difference in change in raw volume	*p* Value	Estimated difference in change in fractional volume (95% CI)	*p* Value
Thalamus	−0.025	0.959	−0.00012	0.658
Caudate	−0.052	0.639	−0.00009	0.458
Putamen	0.581	0.240	0.00045	0.247
Globus pallidus	−0.082	0.617	−0.00009	0.383
Deep gray matter	0.422	0.676	0.00013	0.837

Note: The estimated coefficient is the difference in the annualized change over time in the treated vs. untreated groups. A negative number implies more decrease in volume in the treated subjects.

Discussion

In this pilot study, we showed the utility of an analysis pipeline based on high-resolution 3T MRI for the detection of atrophy of the cerebral DGM over one year in patients with PMS. A statistically significant change in volume was observed in the caudate nucleus and in the total DGM, but all DGM nuclei showed changes in the same direction as the caudate. Considering the use of these metrics in a hypothetical therapeutic treatment trial, with either a single-arm, or two-arm study design, our results indicate feasibility. In contrast, clinical disability (EDSS score) did not significantly change during this one-year observation period. We hypothesize that DGM atrophy may prove efficient as a short-term outcome for proof-of-concept neurotherapeutic trials in PMS.

Our findings are in line with the well-established concept that the DGM are an early and common site of atrophy in patients with MS [17–22, 35–38], which progresses longitudinally [20, 21, 36, 37, 39–41]. The magnitude of decreased DGM volumes we demonstrated in our study during the observation period is inline with these previous studies. There are several potential mechanisms to consider regarding the damage we detected in the DGM during the one-year observation period. First, the caudate and other DGM nuclei are highly metabolically active structures [42] owing to their reciprocal connectivity and core processing role involving nearly all regions of the brain and, in turn, sub-serving a wide range of functions (e.g., cognition, sensation, motor control, visual function, behavior) [43–45]. Thus, given the diffuse and widespread damage, including axonal loss, of WM both in lesions and areas free of overt lesions (normal-appearing WM) in patients with MS [46], the DGM are vulnerable to diaschisis and Wallerian degeneration from damage in the brain and spinal cord [22, 47]. Second, emerging histologic and ultra-high field MRI studies have shown a propensity for the disease process, particularly in PMS, to directly affect the DGM by the presence of lesions [48]. Histologically, these DGM MS lesions are characterized by demyelination, inflammation, microglial and macrophage proliferation, and axonal damage [49]. Third, the DGM are a known site of excessive iron deposition in patients with MS, which might reflect an epiphenomenon (marker) of neurodegeneration, or indeed, may contribute to tissue loss through oxidative stress [50–53]. In addition, the growing body of evidence indicating the entry of activated lymphocytes into the CSF, perhaps through the choroid plexus, raises the possibility of transependymal damage to the DGM, given their proximity to the ventricles [54].

In our study, the use of raw DGM volumes showed higher effect sizes than DGM fractions in the longitudinal demonstration of atrophy. For example, there was a significant loss of volume in raw DGM but not in fractional DGM over one year. These results are in line with our previous study, showing higher scan-rescan reliability for measuring raw vs. fractional DGM [32]. This most likely reflects the lack of precision inherent in the ability of fully automated pipelines to extract the brain contents in the determination of ICV from 3T MRI scans [55]. If such limitations could be overcome in future studies, it is likely that power would increase with normalization of brain volumes, whether it is performed by a residual or proportional calculation [31]. Nonetheless, given that this was a longitudinal study, some normalization was inherent in the use of the patient's baseline volume as a reference for the one-year change.

There are several additional limitations of our study that warrant consideration. A longer observation period and a larger sample size would be useful to extend and confirm these results. As discussed in detail in the Methods section, the clinical management of our patients was heterogeneous and they were not a uniformly untreated (true placebo) group. This may have led to an overestimate of the required sample sizes because the estimated DGM change may not reflect what we would have observed had all subjects been untreated. In addition, each of our sample size calculations used the estimated change and residual standard deviation. But, the limited sample size led to relatively wide confidence intervals for both values. Since the sample size calculations would change if alternative values of these features were used, we encourage researchers designing future studies to assess the sample size requirements for a range of each features based on the reported 95% confidence interval rather than simply using a single value. Given the availability of other fully automated pipelines to measure DGM volumes [56], the effect of the type of pipeline on effect sizes should be considered. Furthermore, a wider breadth of metrics should be compared in PMS for longitudinal sensitivity, vs. DGM volume, such as whole brain [10] and spinal cord atrophy [57]. Future studies should

also examine the effect of lesion filling to refine the DGM segmentations [58]. We examined a mixed sample of two types of progressive forms of MS, to which, collectively, our sample size calculations would apply. However, a subgroup exploratory analysis suggested that the SP group had a lower rate of atrophy than the PP group. Therefore, future studies should also investigate a larger data set of the two groups separately to determine any differential effect sizes for longitudinal change.

Conclusions

Using an automated segmentation pipeline from high-resolution 3T MRI scans, we detected cerebral DGM atrophy over one year in patients with progressive forms of MS. Power calculations based on the magnitude of such atrophy showed promise as a potential outcome measure of neuroprotection in clinical therapeutic trials for this patient population. For example, a trial of a candidate therapy showing a 50% reduced rate of caudate atrophy requires as few as 93 patients for a single-arm study. A therapy showing a 25% reduced rate of caudate atrophy requires 365 patients for a single-arm study. Thus, DGM atrophy may prove efficient as a short-term outcome measure for proof-of-concept neurotherapeutic trials in patients with progressive forms of MS.

Acknowledgments

This work was presented in preliminary form at the 2016 Annual Meeting of the American Academy of Neurology, Vancouver, Canada.

Disclosure statement

The authors have no relevant conflicts of interest to disclose.

ORCID

Rohit Bakshi http://orcid.org/0000-0001-8601-5534

References

1. Filippi M, Rocca MA, Arnold DL, et al. EFNS guideline on the use of neuroimaging in the management of multiple sclerosis. Eur J Neurol 2006;13:313–25.
2. Filippi M, Wolinsky JS, Comi G; CORAL Study Group. Effects of oral glatiramer acetate on clinical and MRI-monitored disease activity in patients with relapsing multiple sclerosis: a multicentre, double-blind, randomised, placebo-controlled study. Lancet Neurol 2006;5:213–20.
3. Zivadinov R, Bakshi R. Role of MRI in multiple sclerosis I: inflammation and lesions. Front Biosci 2004;9:665–83.
4. Bakshi R, Dandamudi VS, Neema M, et al. Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. J Neuroimaging 2005;15:S30–45.
5. Li DK, Held U, Petkau J et al; Sylvia Lawry Centre for MS Research. MRI T2 lesion burden in multiple sclerosis: a plateauing relationship with clinical disability. Neurology 2006;66:1384–9.
6. Ontaneda D, Fox RJ, Chataway J. Clinical trials in progressive multiple sclerosis: lessons learned and future perspectives. Lancet Neurol 2015;14:208–23.
7. Sorensen PS, Blinkenberg M. The potential role for ocrelizumab in the treatment of multiple sclerosis: current evidence and future prospects. Ther Adv Neurol Disord 2016;9:44–52.
8. Vesterinen HM, Connick P Irvine CM, Sena ES, et al. Drug repurposing: a systematic approach to evaluate candidate oral neuroprotective interventions for secondary progressive multiple sclerosis. PLoS One 2015;10:e0117705.
9. Barkhof F, Calabresi PA, Miller DH, et al. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. Nat Rev Neurol 2009;5:256–66.
10. Chataway J, Schuerer N, Alsanousi A, et al. Effect of high-dose simvastatin on brain atrophy and disability in secondary progressive multiple sclerosis (MS-STAT): a randomised, placebo-controlled, phase 2 trial. Lancet 2014;383:2213–21.
11. Dell'Oglio E, Ceccarelli A, Glanz BI, et al. Quantification of global cerebral atrophy in multiple sclerosis from 3T MRI using SPM: the role of misclassification errors. J Neuroimaging 2015;25:191–9.
12. Bakshi R, Neema M, Tauhid S, et al. An expanded composite scale of MRI-defined disease severity in multiple sclerosis: MRDSS2. Neuroreport 2014;25:1156–61.
13. Janardhan V, Bakshi R. Quality of life and its relationship to brain lesions and atrophy on magnetic resonance images in 60 patients with multiple sclerosis. Arch Neurol 2000;57:1485–91.
14. Tauhid S, Chu R, Sasane R, et al. Brain MRI lesions and atrophy are associated with employment status in patients with multiple sclerosis. J Neurol 2015;262:2425–32.
15. Rojas JI, Patrucco L, Míguez J, et al. Brain atrophy in radiologically isolated syndromes. J Neuroimaging 2015;25:68–71.
16. Sanfilipo MP, Benedict RH, Sharma J, et al. The relationship between whole brain volume and disability in multiple sclerosis: a comparison of normalized gray vs. white matter with misclassification correction. Neuroimage 2005;26:1068–77.
17. Ceccarelli A, Jackson JS, Tauhid S, et al. The impact of lesion in-painting and registration methods on voxel-based morphometry in detecting regional cerebral gray matter atrophy in multiple sclerosis. AJNR Am J Neuroradiol 2012;33:1579–85.
18. Bermel RA, Innus MD, Tjoa CW, et al. Selective caudate atrophy in multiple sclerosis: a 3D MRI parcellation study. Neuroreport 2003;14:335–9.
19. Houtchens MK, Benedict RH, Killiany R, et al. Thalamic atrophy and cognition in multiple sclerosis. Neurology 2007;69:1213–23.
20. Modica CM, Bergsland N, Dwyer MG, et al. Cognitive reserve moderates the impact of subcortical gray matter atrophy on neuropsychological status in multiple sclerosis. Mult Scler 2016;22:36–42.
21. Pagani E, Rocca MA, Gallo A, et al. Regional brain atrophy evolves differently in patients with multiple sclerosis

according to clinical phenotype. AJNR Am J Neuroradiol 2005;26:341–6.
22. Ruggieri S, Petracca M, Miller A, et al. Association of deep gray matter damage with cortical and spinal cord degeneration in primary progressive multiple sclerosis. JAMA Neurol 2015;72:1466–74.
23. Chu R, Tauhid S, Glanz BI, et al. Whole brain volume measured from 1.5T versus 3T MRI in healthy subjects and patients with multiple sclerosis. J Neuroimaging 2016;26:62–7.
24. Sicotte NL, Voskuhl RR, Bouvier S, et al. Comparison of multiple sclerosis lesions at 1.5 and 3.0 Tesla. Invest Radiol 2003;38:423–7.
25. Stankiewicz JM, Glanz BI, Healy BC, et al. Brain MRI lesion load at 1.5T and 3T versus clinical status in multiple sclerosis. J Neuroimaging 2011;21:e50–6.
26. Wattjes MP, Harzheim M, Kuhl CK, et al. Does high-field MR imaging have an influence on the classification of patients with clinically isolated syndromes according to current diagnostic MR imaging criteria for multiple sclerosis? AJNR Am J Neuroradiol 2006;27:1794–8.
27. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 2014;83:278–86.
28. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33:1444–52.
29. Fischer JS, Rudick RA, Cutter GR, et al. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. Mult Scler 1999;5:244–50.
30. Gauthier SA, Glanz BI, Mandel M, et al. A model for the comprehensive investigation of a chronic autoimmune disease: the multiple sclerosis CLIMB study. Autoimmun Rev 2006;5:532–6.
31. Sanfilipo MP, Benedict RH, Zivadinov R, et al. Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. Neuroimage 2004;22:1732–43.
32. Chu R, Hurwitz S, Tauhid S, et al. Deep gray matter volume measured from 1.5T vs. 3T MRI in healthy subjects and patients with multiple sclerosis. 2015 meeting of the European Committee on Treatment and Research in Multiple Sclerosis (ECTRIMS), Barcelona. Multiple Sclerosis 2015;23 (suppl):973.
33. Lublin F, Miller DH, Freedman MS, et al; INFORMS study investigators. Oral fingolimod in primary progressive multiple sclerosis (INFORMS): a phase 3, randomised, double-blind, placebo-controlled trial. Lancet 2016;387:1075–84.
34. Steiner D, Arnold DL, Freedman MS, et al. Natalizumab versus placebo in patients with secondary progressive multiple sclerosis (SPMS): results from ASCEND, a multicenter, double-blind, placebo-controlled, randomized phase 3 clinical trial. 68th Annual American Academy of Neurology meeting; 15–21 April 2016; Vancouver, Canada.
35. Batista S, Zivadinov R, Hoogs M, et al. Basal ganglia, thalamus and neocortical atrophy predicting slowed cognitive processing in multiple sclerosis. J Neurol 2012; 259:139–46.
36. Bergsland N, Zivadinov R, Dwyer MG, et al. Localized atrophy of the thalamus and slowed cognitive processing speed in MS patients. Mult Scler 2016;22:1327–3
37. Sepulcre J, Sastre-Garriga J, Cercignani M, et al. Regional gray matter atrophy in early primary progressive multiple sclerosis: a voxel-based morphometry study. Arch Neurol 2006;63:1175–80.
38. 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;59:331–9.
39. Taylor I, Butzkueven H, Litewka L, et al. Serial MRI in multiple sclerosis: a prospective pilot study of lesion load, whole brain volume and thalamic atrophy. J Clin Neurosci 2004;11:153–8.
40. Zivadinov R, Bergsland N, Dolezal O, et al. Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years. AJNR Am J Neuroradiol 2013;34:1931–9.
41. Zivadinov R, Havrdova E, Bergsland N, et al. Thalamic atrophy is associated with development of clinically definite multiple sclerosis. Radiology 2013;268:831–41.
42. Bakshi R, Miletich RS, Kinkel PR, et al. High-resolution fluorodeoxyglucose positron emission tomography shows both global and regional cerebral hypometabolism in multiple sclerosis. J Neuroimaging 1998;8:228–34.
43. Hélie S, Ell SW, Ashby FG. Learning robust cortico-cortical associations with the basal ganglia: an integrative review. Cortex 2015;64:123–35.
44. Ikemoto S, Yang C, Tan A. Basal ganglia circuit loops, dopamine and motivation: a review and enquiry. Behav Brain Res 2015;290:17–31.
45. Silberberg G, Bolam JP. Local and afferent synaptic pathways in the striatal microcircuitry. Curr Opin Neurobiol 2015;33:182–7.
46. Ciccarelli O, Barkhof F, Bodini B, et al. Pathogenesis of multiple sclerosis: insights from molecular and metabolic imaging. Lancet Neurol 2014;13:807–22.
47. Kipp M, Wagenknecht N, Beyer C, et al. Thalamus pathology in multiple sclerosis: from biology to clinical application. Cell Mol Life Sci 2015;72:1127–47.
48. Harrison DM, Oh J, Roy S, et al. Thalamic lesions in multiple sclerosis by 7T MRI: clinical implications and relationship to cortical pathology. Mult Scler 2015;21:1139–50.
49. Vercellino M, Masera S, Lorenzatti M, et al. Demyelination, inflammation, and neurodegeneration in multiple sclerosis deep gray matter. J Neuropathol Exp Neurol 2009;68:489–502.
50. Bergsland N, Tavazzi E, Laganà MM, et al. White matter tract injury is associated with deep gray matter iron deposition in multiple sclerosis. J Neuroimaging 2017;27:107–13.
51. Haider L. Inflammation, iron, energy failure, and oxidative stress in the pathogenesis of multiple sclerosis. Oxid Med Cell Longev 2015; 2015:725370.
52. Quinn MP, Gati JS, Klassen ML, et al. Increased deep gray matter iron is present in clinically isolated syndromes. Mult Scler Relat Disord 2014;3:194–202.
53. Stankiewicz J, Panter SS, Neema M, et al. Iron in chronic brain disorders: imaging and neurotherapeutic implications. Neurotherapeutics 2007;4:371–86.

54. Kivisäkk P, Imitola J, Rasmussen S, et al. Localizing central nervous system immune surveillance: meningeal antigen-presenting cells activate T cells during experimental autoimmune encephalomyelitis. Ann Neurol 2009;65:457–69.
55. Del Re EC, Gao Y, Eckbo R, et al. A new MRI masking technique based on multi-atlas brain segmentation in controls and schizophrenia: a rapid and viable alternative to manual masking. J Neuroimaging 2016;26:28–36.
56. Corrêa DG, Zimmermann N, Tukamoto G, et al. Longitudinal assessment of subcortical gray matter volume, cortical thickness, and white matter integrity in HIV-positive patients. J Magn Reson Imaging 2016;44:1262–9.
57. Valsasina P, Rocca MA, Horsfield MA, et al. A longitudinal MRI study of cervical cord atrophy in multiple sclerosis. J Neurol 2015;262:1622–8.
58. Gelineau-Morel R, Tomassini V, Jenkinson M, Johansen-Berg H, Matthews PM, Palace J. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. Hum Brain Mapp 2012;33:2802–14.