On-line Table 1: Demographics and characteristics ($n = 29$)

Demographic/Characteristic	Mean (SD)
Age (yr)	43.4 (9.7)
Sex	28 women/1 man
Education	15.4 (2.6)
Disease duration (yr)	6.55 (5.5)
EDSS[a]	3.0 (1.5)
MMSE	28.9 (1.5)
Intelligence estimate (WTAR)	106.6 (14.6)
Brief Repeatable Battery	
Selective Reminding Test	
Consecutive long-term retrieval	37.1 (16.3)
Long-term retrieval	48.1 (12.9)
Delayed	8.0 (3.1)
10/36 Spatial Recall	19.9 (5.3)
10/36 Spatial Recall Delayed	6.9 (2.4)
Symbol Digit Modalities Test	46.9 (15.4)
PASAT 2-second	31.0 (11.3)
PASAT 3-second	41.9 (14.4)
Word List Generation	26.9 (10.1)
Flanker interference score[b]	60.8 (39.8)
Stroop interference score[c]	132.8 (69.4)

Note:—WTAR indicates Wechsler Test of Adult Reading; PASAT, Paced Auditory Serial Addition Test.

[a] Using estimate for missing EDSS = 2.5.

[b] For the Flanker task, 1 participant performed below chance. The results are presented with 28 participants.

[c] For the Stroop task, data were lost for 1 participant, and 1 participant performed below chance. The results are presented with 27 participants.

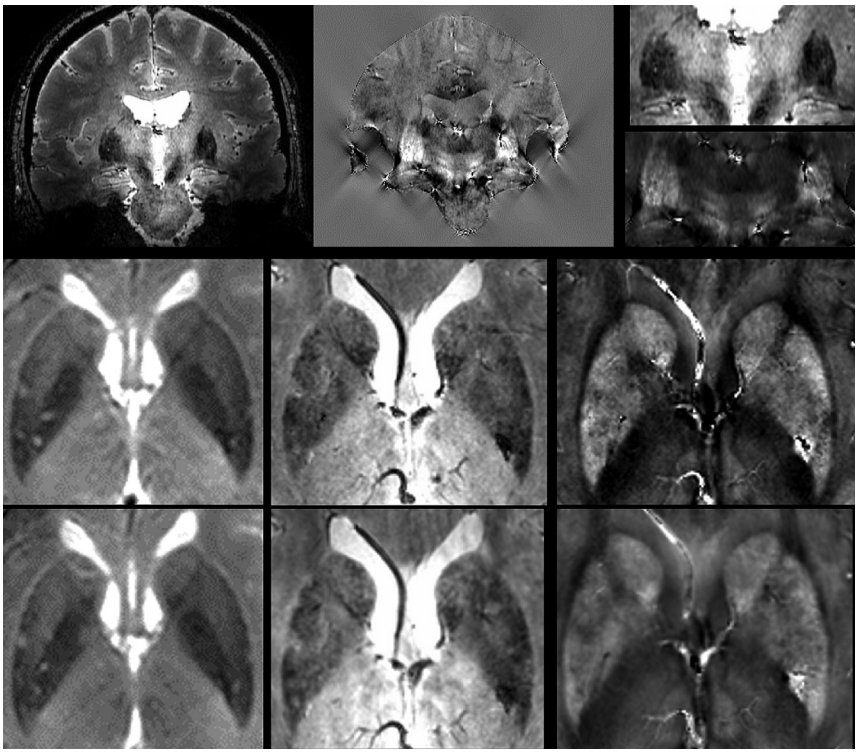
On-line Table 2: Brief Repeatable Battery

Test	Description
Selective Reminding Task	This task is a multitrial memory assessment of encoding and retrieval processes for verbal stimuli in short- and long-term memory. The Selective Reminding Task involved verbal presentation of 12 words, followed by a series of recall trials. After each trial, participants were selectively reminded of words that were not recalled in the previous trial and were directed to recall the entire list of 12 words again. This procedure was repeated for 6 trials or until the participant recalled all 12 words. Delayed recall was administered approximately 11 minutes later and required each participant to freely recall as many words as possible, without cueing. The primary outcome measures were assessed by 3 dependent variables: long-term storage, identified by 2 consecutive recalls of the word; consistent long-term retrieval, referring to a word that is recalled consistently for all subsequent trials in the task; and delayed recall, an uncued free-recall test of any 12 words remembered after an 11-minute delay. The long-term storage score allows the assessment of encoding success, while the consecutive long-term retrieval score is a measure of retrieval success from immediate memory.
10/36 Spatial Recall Task	This is a task of short-term and long-term spatial memory. A checkerboard was placed in front of each participant for 10 seconds with a particular pattern displayed. The participant was asked to replicate the pattern, from memory, on an empty checkerboard. Three total trials of the same pattern were administered or until the participant attained a perfect score. After a 15-minute delay, the participant was asked to recall the pattern, without cueing, to test delayed recall accuracy. Accuracy was the primary outcome variable for immediate and delayed conditions.
Oral Symbol Digit Modalities Test	Participants were presented with a key at the top of the page, including numbers matched to geometric shapes. The bottom portion of the page contained only geometric shapes with blank spaces below. Participants were asked to verbalize the numbers that corresponded to the geometric shapes, as quickly as possible. The experimenter recorded answers for the participant to remove variability associated with upper motor neuron injury. Participants were given 90 seconds to complete as many trials as possible while preserving accuracy. The number of correctly generated responses was the primary outcome measure.
Word List Generation Task	This is a task of verbal fluency and executive functioning. Each participant was provided with a letter of the alphabet and instructed to generate as many words as he or she could think of that began with that letter in 60 seconds. Three trials of differing letters were completed with the number of correctly generated responses across all 3 trials as the primary outcome measure.
PASAT	This task is an assessment of working memory functioning, processing speed, and simple arithmetic. Participants were presented auditorily with digits at the rate of 3 seconds/digit for 1 condition and 2 seconds/digit for the second condition. Participants were instructed to add each number heard to the number previously heard and report the sum out loud. Response accuracy was the primary outcome measure, calculated separately for 2- and 3-second conditions.

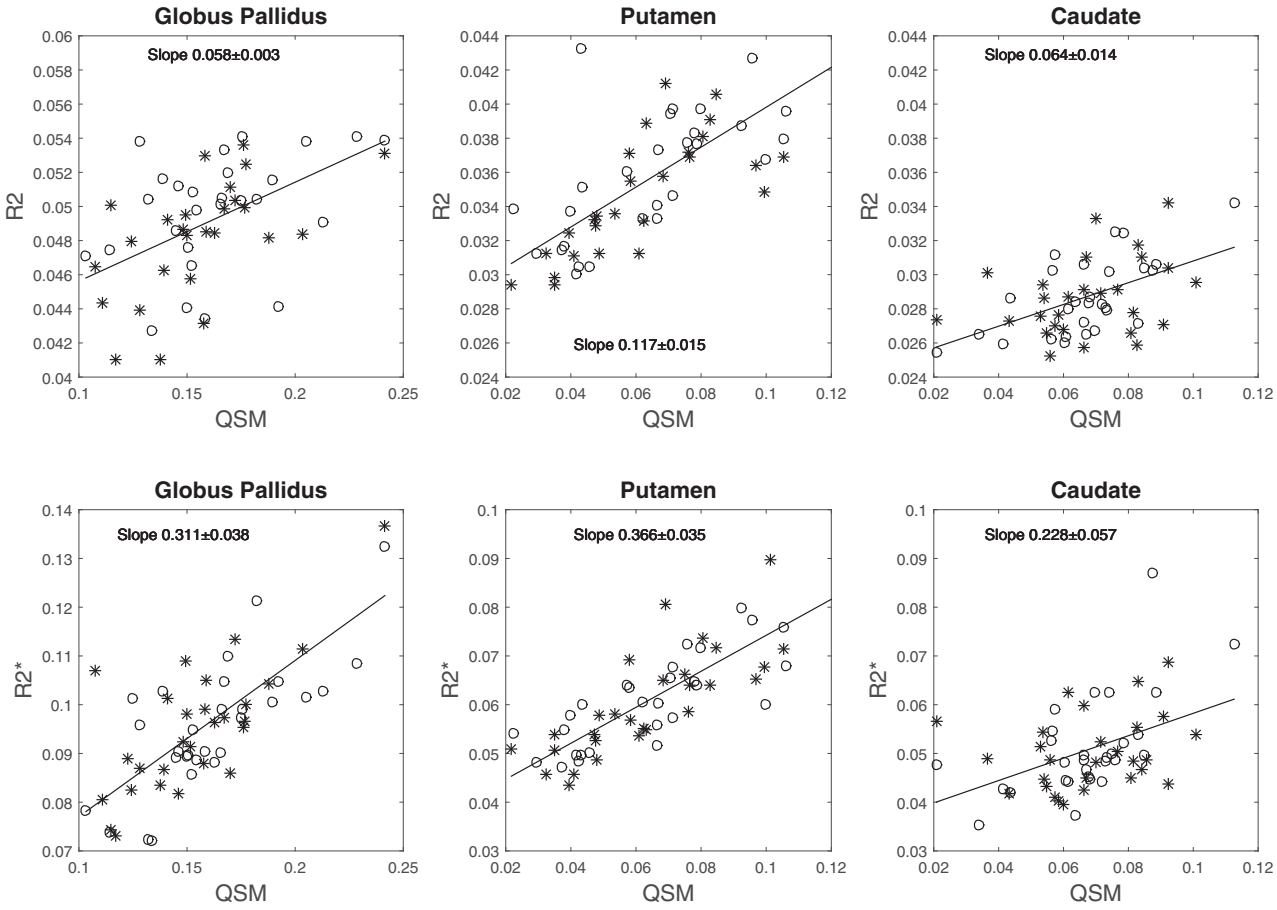
Note:—PASAT indicates Paced Auditory Serial Addition Test.

On-line Table 3: Inhibitory control tasks

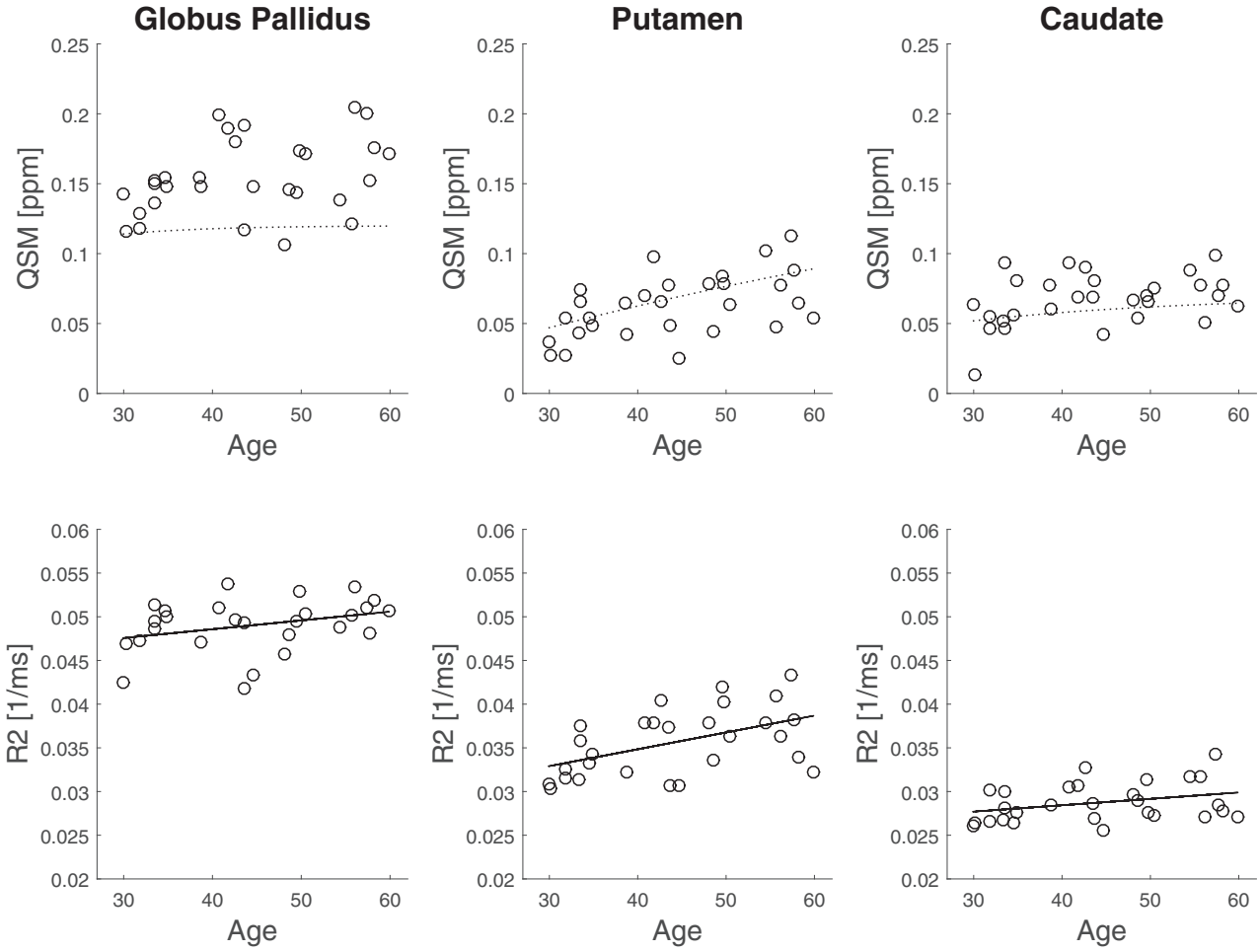
Task	Description
Flanker Task	Participants were presented with a series of 5 arrows on a computer monitor. They were asked to identify the orientation of the central arrow while ignoring the orientation of the peripheral arrows. If the central arrow pointed to the left, the participant was instructed to press the *z* key with their left index finger; when the central arrow pointed to the right, they were instructed to press the right slash key with their right index finger. For half the trials, all 5 arrows pointed in the same direction (congruent condition), and for the other half of the trials, the central arrow pointed in the opposite direction to that of the flanking arrows (incongruent condition). Each stimulus was presented for 1500 ms and an interstimulus interval of 800 ms. There were 100 trials with each condition presented 50 times, congruent or incongruent. Trials were also counterbalanced so that 50% of the arrows pointed left and 50% of the arrows pointed right. The entire task duration lasted approximately 5 minutes. Failures to selectively attend to the task goal (the directionality of the central arrow) and ignore irrelevant information (the flanking arrows) can be identified by increases in reaction time during the incongruent trials compared with performance on congruent trials. "Flanker interference" was calculated by taking the difference between reaction time on congruent trials from incongruent trial performance.
Stroop Task	Participants were administered the Stroop interference task to assess inhibitory control. They were instructed to respond to the color of the ink in which the stimulus was printed rather than the semantic meaning of the word. Participants responded to 4 colors: red, green, blue, and yellow using both their right and left index fingers placed on the *x*, *c*, *n*, and *m* keys, respectively. We had 3 main conditions in this task: congruent (when the color in which the word was printed was same as the semantic meaning of the word, such as "RED" displayed in red ink); neutral (when the word displayed on screen was not a color word but a more neutral word, such as "LAMP" displayed in red ink); and an incongruent condition (when there was conflict between the ink color in which the word was printed and the semantic meaning of the word, such as "BLUE" in red ink). A total of 216 trials, 72 for each of the 3 conditions, were presented. The stimuli were presented to each participant for 2000 ms per trial with an interstimulus interval of 1000 ms. The main outcome variable of interest in this task was "Stroop interference," calculated by subtracting the reaction time for congruent trials from the reaction time of incongruent trials.



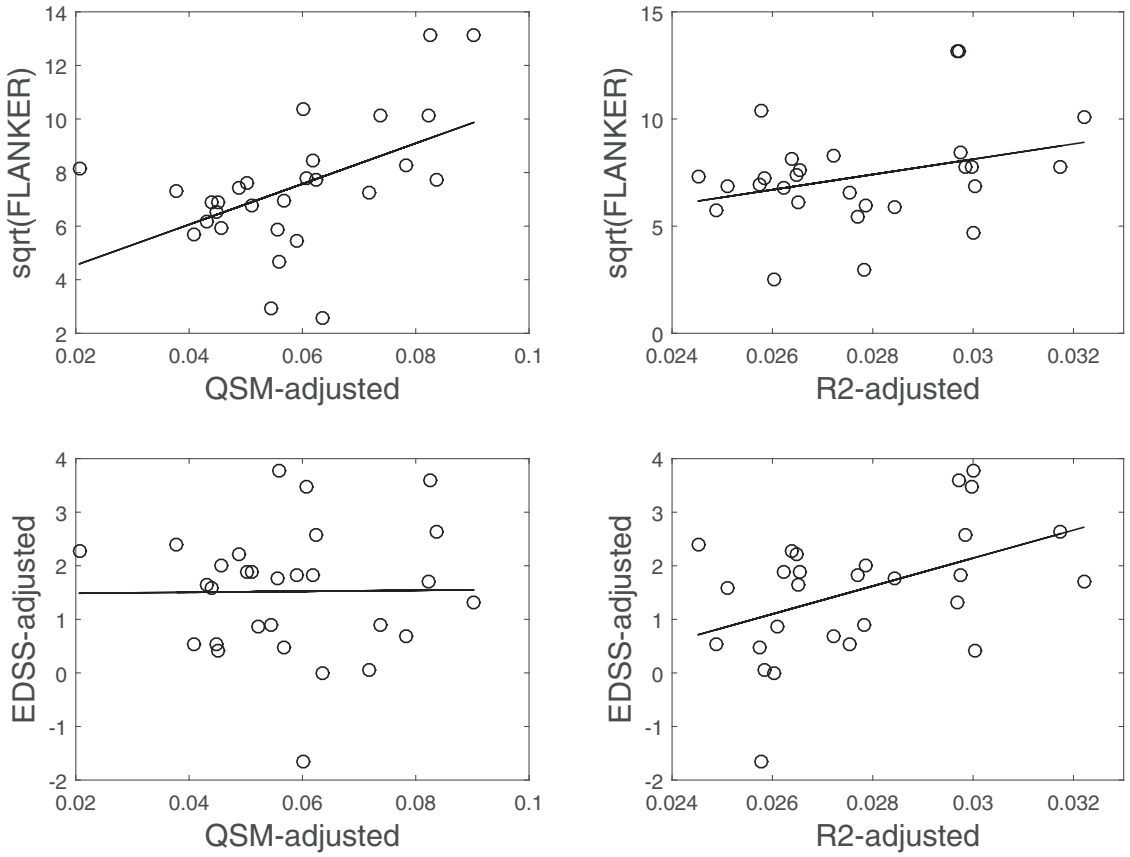
ON-LINE FIG 1. Coronal reformatted T2*WI and QSM (*top row*) and axial magnified T2/T2*-weighted and QSM images of a 57-year-old woman with MS (EDSS = 3; duration, 4.6 years) at baseline (*top*, *middle*) and 12 months later (*bottom*). The coronal reformatted QSM image shows some streaking artifacts from QSM processing, but the BG region is minimally affected. Several high-iron-content regions are seen within the GP, putamen, and the posteromedial tips of the caudate. They are seen dark on T2/T2*WI and bright on QSM. This is persistent with time. Overall iron content is high throughout the putamen (ie, near-isointense to GP); some medium dark regions on QSM are gray on T2*WI and may be calcifications.



ON-LINE FIG 2. Relations between R2 and QSM (*top*) and R2* and QSM (*bottom*) for the GP (*left*), putamen (*middle*), and caudate (*right*). Data for right (*circles*) and left (*stars*) ROIs are included. Slopes for R2 versus QSM are significantly different among structures. Slopes for R2* versus QSM are not significantly different. Note that slopes and correlation coefficients for the caudate are the lowest.



ON-LINE FIG 3. Change of QSM (*top*) and R2 (*bottom*) with age. The *dotted lines* in the QSM data were computed from published analytic expressions for QSM change in healthy aging.[20] Age-related QSM changes in patients with MS follow trends similar to those in normal aging but are greater in the GP and caudate. Caution is required with this comparison because there may be shifts due to differences in QSM processing. The *solid lines* in the R2 data represent linear regression fits.



ON-LINE FIG 4. The relations between cognitive and clinical measures and MR imaging metrics for the caudate. The age dependence of QSM and R2 were removed by subtracting the linear fit line of QSM/R2 versus age from the right/left averaged experimental data. For plotting purposes, the offset was adjusted so that the resultant adjusted QSM and R2 overlapped at 30 years of age. This offset has no effect on subsequent analysis. The analogous EDSS score was adjusted for its dependence on disease duration (square root of duration was used because duration was not normally distributed). The adjusted residual QSM and R2 were then correlated to the square root of the Flanker task interference time (*top*). Square root was used because the Flanker times were not normally distributed. We found a significant correlation between the Flanker and caudate QSM ($r = 0.51$, $P = .006$) but not R2 ($r = 0.31$, $P = .11$). Residual age-adjusted QSM and R2 are also correlated with the duration-adjusted EDSS (*bottom*). EDSS correlates with R2 ($r = 0.46$, $P = .014$) but not QSM ($r = 0.0$, $P = .95$). Shown are plots of the adjusted data with *lines* from linear regression.