HHS Public Access

Author manuscript

Ann Neurol. Author manuscript; available in PMC 2019 January 03.

Published in final edited form as:

Ann Neurol. 2018 February ; 83(2): 223–234. doi:10.1002/ana.25150.

Thalamic Atrophy in Multiple Sclerosis: A Magnetic Resonance Imaging Marker of Neurodegeneration throughout Disease

Christina J. Azevedo, MD, MPH[1], **Steven Y. Cen, PhD**[1], **Sankalpa Khadka, BS**[2], **Shuang Liu, PhD**[2], **John Kornak, PhD**[3], **Yonggang Shi, PhD**[1], **Ling Zheng, PhD**[1], **Stephen L. Hauser, MD**[4], and **Daniel Pelletier, MD**[1]

[1]Department of Neurology, University of Southern California, Los Angeles, CA;

[2]Department of Neurology, Yale University, New Haven, CT;

[3]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA;

[4]Department of Neurology, University of California, San Francisco, San Francisco, CA

Abstract

Objective: Thalamic volume is a candidate magnetic resonance imaging (MRI)-based marker associated with neurodegeneration to hasten development of neuroprotective treatments. Our objective is to describe the longitudinal evolution of thalamic atrophy in MS and normal aging, and to estimate sample sizes for study design.

Methods: Six hundred one subjects (2,632 MRI scans) were analyzed. Five hundred twenty subjects with relapse-onset MS (clinically isolated syndrome, n = 90; relapsing-remitting MS, n = 392; secondary progressive MS, n = 38) underwent annual standardized 3T MRI scans for an average of 4.1 years, including a 1mm 3-dimensional T1-weighted sequence (3DT1; 2,485 MRI scans). Eighty-one healthy controls (HC) were scanned longitudinally on the same scanner using the same protocol (147 MRI scans). 3DT1s were processed using FreeSurfer's longitudinal pipeline after lesion inpainting. Rates of normalized thalamic volume loss in MS and HC were compared in linear mixed effects models. Simulation-based sample size calculations were performed incorporating the rate of atrophy in HC.

Results: Thalamic volume declined significantly faster in MS subjects compared to HC, with an estimated decline of −0.71% per year (95% confidence interval [CI] =−0.77% to −0.64%) in MS subjects and −0.28% per year (95% CI = −0.58% to 0.02%) in HC (p for difference = 0.007). The

Address correspondence to Dr Pelletier, Neuroimmunology Division, Department of Neurology and USC Multiple Sclerosis Center, Keck Schoolof Medicine of the University of Southern California, 1540 Alcazar Street, Suite 206, Los Angeles, CA 90033. Daniel.Pelletier@usc.edu.

Author Contributions

D.P. and C.J.A. were involved in the study concept and design. All authors were involved in data acquisition and analysis. D.P., C.J.A., S.Y.C., and L.Z. were involved in drafting the manuscript and figures.

Potential Conflicts of Interest

D.P. has received consulting and/or speaking fees from Biogen, and grant/research support from Biogen. C.J.A. has received consulting fees from Biogen. Biogen cosponsored the acquisition of the MRI scans used in these studies; however, the authors' work for Biogen was not related to this study, and no Biogen products were used in the study, although some subjects may have been taking such products as part of their routine clinical care. Similarly, although GlaxoSmithKline cosponsored MRI data acquisition, their products were not used in this study, although some subjects may have been taking their products as part of their routine clinical care.

rate of decline was consistent throughout the MS disease duration and across MS clinical subtypes. Eighty or 100 subjects per arm ($\alpha = 0.1$ or 0.05, respectively) would be needed to detect the maximal effect size with 80% power in a 24-month study.

Interpretation: Thalamic atrophy occurs early and consistently throughout MS. Preliminary sample size calculations appear feasible, adding to its appeal as an MRI marker associated with neurodegeneration.

Although difficult to define precisely, neurodegeneration is a fundamental aspect of the pathology of multiple sclerosis (MS). Within inflammatory-mediated white matter demyelinating lesions, a varying degree of axonal transection may occur[1]; however, although most demyelinated axons will survive the acute injury, they do not survive chronic states of demyelination.[2] There are several mechanisms by which white and gray matter degeneration may occur in MS. Iron accumulation,[3] microglial activation,[4] mitochondrial dysfunction,[5] and glutamate excitotoxicity[6] may all lead to a final common pathway of oxidative stress that, when sustained chronically, will overwhelm cellular compensatory mechanisms and ultimately lead to neuronal cell death. Although the dynamics of neuroaxonal degeneration are not well appreciated by crosssectional histopathological studies and likely vary between patients,[2] once neurodegeneration begins, it should typically be a relentless and progressive process over time. Clinically, neuroaxonal loss is thought to be the major pathologic substrate of irreversible physical and cognitive disability in MS.[7] As such, understanding and preventing neurodegeneration is currently a priority in the field.

As a major relay nucleus with extensive cortical and subcortical anatomic connections, the thalamus is a critically important location in MS. Axonal transection within white matter lesions[1] causes disconnections in the tracts projecting into and out of the thalamus (eg, thalamocortical, opticothalamic, spinothalamic), and is likely a major contributor to the reduction in thalamic volume seen on histopathology.[8] The observation that neuronal density is reduced in nonlesional thalamic tissue[8] supports this hypothesis. Thalamic iron deposition[3] and thalamic demyelinating lesions[9] may also contribute to subsequent thalamic atrophy. Thus, thalamic atrophy may reflect the net accumulation of MS-related damage throughout the central nervous system (CNS), making the thalamus a sensitive and appealing location to study neurodegeneration in MS.

Magnetic resonance imaging (MRI) has corroborated many of these histopathologic findings and provides an opportunity to study thalamic pathology in vivo and longitudinally. Early work using MRI and ^{1}H-MR spectroscopy demonstrated a reduction in both thalamic N-acetylaspartate concentration and thalamic volume in relapsing-remitting MS (RRMS), suggesting neuroaxonal loss/dysfunction as the substrate for thalamic atrophy.[10] The presence of thalamic atrophy has been reported cross-sectionally in early CNS demyelinating disease, including clinically isolated syndrome (CIS) at presentation,[11] pediatric MS,[12] and radiologically isolated syndrome.[13] Moreover, nonconventional imaging approaches have been used to study mechanisms of thalamic atrophy in vivo; data using diffusion tensor imaging tractog-raphy suggest that white matter lesions play a central role in thalamic atrophy, and susceptibility-based imaging suggests that iron deposition likely

contributes.[15,16] In addition, work at high field (7T) has found a correlation between periventricular thalamic lesion burden and thalamic atrophy.[17]

Because thalamic atrophy is present early in disease, reflects several aspects of MS pathology including gray matter injury, and correlates well with physical[18] and cognitive[19] impairment, thalamic volume has been proposed as a potentially attractive MRI metric associated with neurodegenerative features of MS.[20] Developing such a metric is a pressing need in the field, which currently lacks a robust, validated MRI metric associated with neurodegeneration that could be less confounded by tissue fluid dynamics. The objective of this work is to describe the longitudinal evolution of thalamic atrophy at the group level, with the hypothesis that thalamic volume declines over time in MS across the duration of the disease. We examine this question in a unique dataset from a large, well-characterized, prospective cohort of MS patients of all disease stages and subtypes in whom MRI was obtained in a standardized fashion. If found, the presence of thalamic volume decline across the spectrum of disease duration and/or clinical phenotypes would provide the rationale to perform preliminary power calculations using thalamic volume as a primary endpoint for phase 2 proof-of-concept studies.

Subjects and Methods

Research Participants

From January 2005 through December 2010, all Caucasian patients age 18 to 65 years who fulfilled 2005 McDonald criteria[21] for CIS/MS were offered enrollment into a large, prospective phe-notype-genotype biomarker study (the EPIC study) at the University of California, San Francisco (UCSF) MS Center. As described previously,[22] > 500 CIS/MS subjects were followed longitudinally with annual standardized clinical visits and brain MRI scans for an average of 5 years each. All EPIC subjects with relapsing-from-onset forms of MS[23] were included in this analysis.

Healthy controls were recruited separately. Brain MRIs were performed serially on the same scanner using the same imaging sequences as MS subjects. Follow-up brain MRIs were scheduled according to healthy control availability; as such, the interval between MRI scans was not standardized in healthy controls. Healthy controls were selected for inclusion in this analysis to match the age and gender distributions of the MS subjects.

MRI Acquisition

All brain MRI scans (2,632 MRI scans) were acquired using an 8-channel phased array coil in reception and a body coil in transmission on a 3T GE Excite scanner (GE Medical Systems, Milwaukee, WI) that did not undergo hardware upgrades during the 5-year study period. A 3-dimensional (3D), T1-weighted, 1mm isotropic, volumetric, inversion recovery spoiled gradient echo (IRSPGR) sequence ($1 \times 1 \times 1mm^3$, 180 slices) was acquired at each time point (echo time/repetition time/inversion time = 2/7/400 milliseconds, flip angle = 15 °, $256 \times 256 \times 180$ matrix, $240 \times 240 \times 180mm^3$ field of view, number of excitations = 1).

MRI Postprocessing

Images were processed using a semiautomated, custom-made, in-house postprocessing pipeline in which the IRSPGR was the input image. In MS subjects, white matter lesions were manually segmented from the baseline IRSPGR and a lesion mask was derived. Using the baseline lesion mask and after coregistration of follow-up images, subsequent time points were processed using an in-house automated lesion segmentation pipeline that generates lesion masks at each time point for a given subject ($\kappa = 0.9$ compared to manual segmentation, unpublished data). Lesion masks were used to inpaint[24] MS lesions at each time point prior to submitting the IRSPGR images to FreeSur-fer's longitudinal processing stream (v5.3; surfer.nmr.mgh.har-vard.edu/fswiki/LongitudinalProcessing).[25]

Briefly, to reduce postprocessing variability and increase statistical power, FreeSurfer's longitudinal pipeline is designed to be unbiased with respect to any given time point because each time point is treated identically, as opposed to methods that may smooth or resample subsequent time points but not the baseline image, for example. FreeSurfer accomplishes this by first processing each time point cross-sectionally, then iteratively registering all time points to the median to create a within-subject template, also referred to as the subject "base." The "base" can be viewed as an "initial guess" for surface reconstruction and subcortical segmentation. It is used to initialize several steps in the longitudinal run, such as intensity normalization, atlas registration, subcortical segmentation, and surface reconstruction. The use of a subject-specific template has been shown to reduce the random variation in FreeSurfer's highly complex processing procedure, thereby improving the sensitivity of the longitudinal analysis to detect small changes in cortical thickness or subcortical volumes.[25]

For this analysis, all FreeSurfer output was reviewed by experienced MRI readers/postprocessors (C.J.A., S.K., S.L.) and manually corrected as needed using methods within FreeSurfer such as control points, white matter edits, and dura edits. Individual cases were rerun as needed until no further improvement was achieved with additional edits. Left and right thalamic volumes were extracted directly from FreeSurfer's final output and added together to create a total thalamic volume at each time point. Thalamic volume was normalized for head size using the total intracranial volume, which was also obtained directly from Free-Surfer; the resulting normalized thalamic volumes are unitless.

To corroborate the robustness of our findings, all 2,632 IRSPGRs from MS and healthy control subjects were processed with FIRST, a separate image-processing method with a different segmentation strategy (FMRIB Integrated Registration and Segmentation Tool, v5.0, Oxford, UK; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIRST).[26] Briefly, FIRST constructs shape/appearance models of subcortical structures using a dataset of manually segmented images to create labels that are parameterized as surface meshes and modeled as a point distribution function. These labels provide training information for segmentation of input images using a Bayesian approach. As opposed to Free-Surfer's longitudinal pipeline, FIRST provides a cross-sectional measurement of thalamic volume at each time point.[26]

Statistical Analysis

AGREEMENT BETWEEN THALAMIC VOLUME MEASUREMENTS FROM FREESURFER VERSUS FIRST.—To estimate the agreement of the two image-processing methods, intraclass correlation was calculated between thalamic volume measurements from FreeSurfer versus those from FIRST at each time point in all subjects.

ESTIMATION OF THE RATE OF THALAMIC VOLUME DECLINE.—All statistical analysis was done in SAS (v9.4; www.sas.com). The rate of normalized thalamic volume loss was estimated in MS subjects and healthy controls in a linear mixed effects model[27] that can be expressed as follows: $Y_{ij} = \beta_{00} + \beta_{10}\,(Time)_{ij} + \beta_{01}(MS) + \beta_{11}(MS) \times (Time)_{ij} + b_{0j} + b_{1j}(Time)_{ij} + \varepsilon_{ij}$.[28] This model includes time from study entry in years (*Time*), a binary variable for disease status *(MS)* as a fixed effect, an interaction between disease status and time (*MS* × *Time*), random effects for both intercept and slope (b_{0j} and b_{1j} respectively), and $\varepsilon_{ij} \sim iid\ N(0, \sigma^2)$. Age at study entry was tested as a covariate; the addition of age at study entry and its interaction with time to the mixed model did not significantly alter the rate of normalized thalamic volume decline. As such, age was not retained as a covariate in the models presented. Thalamic volume decline was modeled both as an absolute decline and as a percentage change from baseline; the latter was performed to express the results in terms that are directly comparable to prior work in other regions (eg, whole brain volume).

Additional linear mixed effects models were used to estimate the rate of normalized thalamic volume loss in each 5-year epoch of disease duration at study entry and in each clinical subtype at study entry (CIS, RRMS, secondary progressive MS [SPMS]).[23] In each mixed effects model, random (subject-specific) effects for both intercept and slope were included to allow estimation of both the between- and within-subject variability in thalamic volume. In addition, the impact of cumulative exposure to disease-modifying therapy (DMT) on the rate of thalamic volume decline was estimated. Lifetime cumulative DMT exposure prior to each MRI scan date was calculated, including any DMT exposure that occurred either prestudy or on-study. We modeled DMT exposure in linear mixed effects models using 2 approaches: (i) as the cumulative lifetime exposure at baseline for each subject (ie, at study entry) and (2) as a time-varying covariate.

CORRELATION WITH CLINICAL ENDPOINTS AND WHOLE BRAIN VOLUME.—Spearman correlation was used to measure the association between baseline thalamic volume and clinical metrics, including the Expanded Disability Status Scale (EDSS),[29] the Multiple Sclerosis Functional Composite (MSFC),[30] and each of the 3 MSFC components: Paced Auditory Serial Addition Test (PASAT), 9-Hole Peg Test, and Timed 25-Foot Walk. We also examined the correlation between percentage change in thalamic volume over i and 2 years and percentage change in whole brain volume, measured by Structural Image Evaluation Using Normalization of Atrophy (SIENA),[31] over 2 and 5 years, using Spearman correlation.

SAMPLE SIZE CALCULATIONS.—Using parameter estimates directly from mixed effects models, simulation-based sample size calculations[32] were performed to estimate the number of subjects that would be needed per arm to detect a given reduction in the rate of

thalamic atrophy in randomized controlled trials of 12- and 24-month duration. Briefly, unlike commonly used cross-sectional formulas for power estimation, the simulation-based calculations use an empirical distribution with a complex variance covariance structure. In this particular study, we included the slope estimates (ie, βs), the variance−covariance matrix for random effects, and the error variance estimated directly in a linear mixed effects model. Assuming these parameters reflect the true underlying distribution of the data and holding them constant, the procedure takes N random independent replicates under each predetermined sample size (in our case, 5,000 replicates at total sample sizes of 80−280 by increments of 20). The procedure fits the model to each replicate and determines the proportion of replicates that lead to rejection of the null hypothesis under each sample size scenario. The power curve is then produced with points interpolated using penalized B-spline. Because of the difference in intervals between MRI scans for the MS subjects and healthy controls, we used variance parameters only from the MS subjects in the simulation procedure.

Sample size calculations were performed in all MS subjects together, and in treated subjects only, to consider different clinical trial scenarios. We assumed a "maximal effect size," defined as the difference between thalamic volume decline in MS subjects compared with healthy controls; in other words, the maximal effect size would refer to a hypothetical treatment that could slow the rate of thalamic volume decline in MS to that of normal aging. We performed sample size calculations using an alpha of both 0.05 and 0.1; some have advocated an alpha of 0.1 in phase 2 clinical trials.[33] Because MS subjects in the EPIC study underwent MRIs at fixed annual intervals, the 12-month study design was assumed to include 2 MRI time points (0 and 12 months), and the 24-month study design included 3 MRI time points (0, 12, and 24 months).

Results

Subject Demographics

Table 1 includes demographics, baseline characteristics, and number of MRI time points for all 601 subjects. Five hundred twenty subjects with relapsing-from-onset forms of MS were included (CIS, n = 90; RRMS, n = 392; SPMS, n = 38), as well as 81 healthy controls. As expected based on matching, there was no statistically significant difference in age or gender between MS subjects and healthy controls.

Intraclass Correlation of Thalamic Volume Measured with FreeSurfer versus FIRST

The intraclass correlation between thalamic volumes measured with FreeSurfer versus those measured with FIRST was 0.91 (95% confidence interval [CI] = 0.90 to 0.92) overall and >0.90 at each time point, indicating a very high level of agreement between thalamic volume measurements using these two methods.

All MS Subjects versus Healthy Controls

Figure 1 shows estimated normalized thalamic volume over time in all MS subjects and healthy controls (right and left panels, respectively). MS subjects had a statistically significantly lower baseline normalized thalamic volume (ie, lower intercept) compared to

healthy controls (0.925 ± 0.01 vs 1.01 ± 0.01, $p < 0.001$). The rate of thalamic volume loss (percentage change from baseline) was −0.28% per year (95% CI = −0.58% to 0.02%) in healthy controls and −0.71% per year (95% CI = −0.77% to −0.64%) in MS subjects. The rate of thalamic volume decline was significantly faster in MS subjects, with a parameter estimate of an additional −0.42% thalamic volume decline per year (95% CI = −0.73% to −0.11%, $p = 0.007$).

MS Subjects by Disease Duration at Study Entry

MS subjects were divided into 5-year epochs of disease duration at study entry, and rates of normalized thalamic volume loss were estimated in each epoch (Fig 2). A similar estimated percentage of thalamic volume decline was seen across all epochs; estimated rates were −0.68% per year (95% CI = −0.78% to −0.58%) in epoch 1, −0.75% per year (95% CI = −0.89% to −0.61%) in epoch 2, −0.68% per year (95% CI = −0.85% to −0.51%) in epoch 3, and −0.72% per year (95% CI = −0.86% to −0.57%) in epoch 4. The intercepts progressively decreased in each epoch as disease duration increased (0.963 in epoch 1, 0.927 in epoch 2, 0.902 in epoch 3, and 0.855 in epoch 4), consistent with a linear decline in thalamic volume throughout the duration of the disease. To further demonstrate that thalamic volume decline was consistent across the disease, disease duration at study entry was tested as a predictor in linear mixed effects models. When disease duration was modeled categorically, the time × disease duration interaction test was not statistically significant (3 *df*, $p = 0.88$). When modeled as a continuous predictor, disease duration at study entry did not significantly predict thalamic volume decline (0% additional decline for each 5-year increase in disease duration, 95% CI = −0.04% to 0.04%, $p = 0.99$). Finally, in the model that included disease duration and the interaction between disease duration and time, the marginal effect of disease duration was estimated to be −0.04 ± 0.01 per year ($p < 0.001$), suggesting that there is a statistically linear decline in the intercepts of each epoch.

MS Subjects by Clinical Subtype at Study Entry

Rates of normalized thalamic volume loss were estimated in each clinical disease subtype at study entry (Fig 3). Normalized thalamic volume declined in each subtype of MS, with estimated rates of −0.58% (95% CI = −0.74% to −0.43%), −0.74% (95% CI = −0.82% to −0.67%), and −0.59% (95% CI = −0.83% to −0.35%) per year in CIS, RRMS, and SPMS, respectively. The intercepts progressively decreased as disease subtype worsened (0.969 in CIS, 0.923 in RRMS, and 0.841 in SPMS). Of the 392 subjects whose clinical subtype was labeled as “relapsing-remitting” at baseline, only 21 were described as “secondary progressive” by the treating physician by the end of the 5-year study (5.4%). Of the 90 who were classified as CIS at study entry, 47 were reclassified to RRMS or SPMS by the treating physician during the study period (52.2%). Similar to the disease duration analysis, when clinical subtype was tested as a categorical predictor of thalamic volume decline, the time × clinical subtype interaction test was not statistically significant (2 *df*, $p = 0.12$), suggesting a consistent decline in thalamic volume across the spectrum of MS clinical severity.

Effect of DMT on Thalamic Volume Decline

When modeled as a fixed effect, we found that lifetime cumulative exposure to DMT at study entry did not statistically significantly impact the rate of thalamic volume decline

$\beta_{DMT \times time}$ = 0.005% (95% CI = −0.029% to 0.040%, p = 0.77) for each additional year of exposure. When modeled as a time-varying covariate, we found that higher lifetime cumulative DMT exposure was associated with a significantly slower rate of thalamic volume decline $\beta_{\mathrm{DMT} \times time}$ = 0.041% (95% CI = 0.021% to 0.061%, p < 0.001).

Clinical Relevance: Correlation with Clinical Endpoints and Whole Brain Volume

The correlation between thalamic volume and clinical endpoints is given in Table 2. Each correlation is statistically significant (although not particularly high) and in the expected direction; as thalamic volume decreases, MSFC and PASAT also decrease, whereas EDSS, 9-Hole Peg Test, and Timed 25-Foot Walk increase. Statistically significant correlations between percentage change in thalamic volume and percentage change in whole brain volume as measured by SIENA were also seen (Table 3). The correlation with whole brain volume was generally larger when considered in terms of percentage change in thalamic volume over 2 years rather than 1 year.

Clinical Relevance: Simulation-Based Sample Size Estimations

The “maximal effect size” concept is illustrated in Figure 4A. Figure 4B shows the power function in CIS and RRMS patients assuming α = 0.1. To achieve 80% power, we estimated that 80 subjects would be needed per arm to detect the maximal effect size in a 24-month study, or 118 per arm in a 12-month study. Assuming α = 0.05 in CIS and RRMS, to achieve 80% power, 100 subjects would be needed per arm to detect the maximal effect size in a 24-month study, and >140 subjects per arm in a 12-month study (see Fig 4C). In all patients (CIS, RRMS, and SPMS), assuming α = 0.1, 85 subjects would be needed per arm to detect the maximal effect size in a 24-month study, or 128 per arm in a 12-month study. Assuming α = 0.05 in all subjects, 109 per arm would be needed in a 24-month study, and >140 in a 12-month study.

Discussion

This work describes the annual rate of thalamic atrophy in MS in a large, well-characterized cohort of subjects with standardized 3D volumetric imaging, robust image processing, and advanced statistical analysis. Our findings demonstrate that thalamic volume declines over time in MS, at a rate that is significantly faster than that seen in healthy controls. Our estimated rates of thalamic volume decline are consistent with what is reported in the literature for whole brain atrophy in both MS and in healthy controls (−0.71% and −0.28% per year, respectively).[34,35] Furthermore, at the group level, our data suggest that the rate of thalamic volume decline is consistent across MS clinical subtypes, and perhaps more importantly, throughout the duration of the disease. This consistent rate of decline (ie, sensitivity to change over time), in combination with its sensitivity in early disease phases and reflection of multiple aspects of the disease pathobiology, suggests that thalamic volume may be appropriate for use as a primary MRI endpoint in a neu-roprotective trial.

There is substantial biologic plausibility to our findings. Using diffusion tensor imaging tractography, a mechanistic relationship between white matter lesions within thalamocortical projections and thalamic atrophy has been shown in vivo in CIS subjects,[14] suggesting that

axonal transection occurring within white matter MS lesions[1] in tracts projecting into the thalamus may cause upstream and downstream degeneration and a subsequent reduction in thalamic volume. The presence of thalamocortical tract-specific patterns of pathology, specifically an association between myelin loss in nonlesional white matter tracts and neurodegeneration in connected cortical and thalamic gray matter regions, has been reported in postmortem his-topathology and supports the presence of downstream, upstream, or transsynaptic degeneration in MS.[36] In addition, oxidative stress associated with thalamic iron deposition may contribute to thalamic volume loss[15,16] via a vicious cycle in which inflammatory demyelination attracts iron-laden macrophages, leading to iron overload, resulting in free radical generation, which in turn reduces cellular energy production and causes hypoxia, progressive axonal dysfunction, further inflammation, and more iron deposition. Thalamic demyelinating lesions may contribute to thalamic volume loss as well.[9,17] Thus, thalamic volume measurements on MRI may reflect the overall consequences of the disease, which could explain the presence of thalamic volume loss in early MS and its persistent decline throughout the disease duration.

If thalamic atrophy is truly an early in vivo measure of progressive tissue loss, then these data would support the notion that there is not a single, biological transition point in the disease at which patients “convert” from RRMS to SPMS; rather, disease progression/neurodegeneration starts from the very beginning of the disease and proceeds relentlessly. This concept is not reflected in our current clinical definitions, which necessarily assume a transition point. Although valuable, our clinical definitions are not necessarily reflective of, and in some ways may be quite far away from, the underlying biology of the disease. This fundamental disconnection between clinical versus biological characterizations of MS has important implications for MS clinical research and clinical trial design. One could argue that a biological characterization of the disease is perhaps preferable, and is the definition of a biomarker. A robust, sensitive, and specific biomarker associated with neurodegeneration has been lacking and is necessary for the field of MS to move forward in stopping disease progression.

We performed an initial investigation into the clinical relevance of thalamic volume, and found statistically significant correlations with EDSS, MSFC, and all 3 MSFC components. Each correlation was in the a priori expected direction,[19,37,38] but of modest strength. In a cross-sectional analysis, this might be expected. In particular, when considering atrophy-based markers, one must consider the possibility that the MRI and clinical variables may not be changing concurrently; in other words, it is plausible that tissue loss on MRI precedes the clinical consequences of those changes by many years. As such, longitudinal analyses of the clinical relevance of thalamic atrophy at the individual level will be an important future direction but are beyond the objective of this work. In addition, we provide estimated correlations with whole brain volume measured by SIENA, a commonly used MRI metric associated with neurodegeneration, which are in the expected direction. A full analysis comparing thalamic volume and whole brain volume is again outside the scope of the current work, but will be of great interest in future analyses.

As a sensitivity analysis, we investigated the effects of DMT on the rate of thalamic atrophy. Because this observational cohort study occurred between 2005 and 2010, nearly all of the

subjects exposed to DMT were on first-generation, injectable DMTs. When lifetime cumulative DMT exposure at study entry was modeled as a simple baseline effect, there was no statistically significant effect of DMT on the rate of thalamic atrophy. Interestingly, when modeled as a time-varying covariate, longer DMT exposure was associated with significantly slower thalamic volume decline. It is important to understand the difference in interpretation of these two modeling approaches. The focus of this analysis is to describe an MRI outcome measure that can be used in clinical trials to make between-group comparisons, which is how a baseline effect is interpreted. A time-varying covariate, conversely, is interpreted as the change in thalamic atrophy within an individual, according to how DMT status is changing over time. As such, the baseline effect approach is perhaps more appropriate to the current work, because in a drug trial DMT exposure is typically controlled and would not be modeled as a time-varying covariate. However, the time-varying approach is also interesting in its own right, and should be the focus of a future analysis. It is important to note that, as an observational cohort study rather than a randomized controlled trial, this study was not designed or powered to see a between-treatment-group effect on any endpoint, including thalamic atrophy. The absence of a treatment effect when modeled as a between-group comparison (ie, fixed effect) should not suggest that thalamic atrophy is not modifiable with treatment; a recently presented post hoc analysis from a randomized controlled trial of a newer DMT has shown a treatment effect on thalamic atrophy.[39]

Because of its appeal as an MRI biomarker, we performed simulation-based power calculations using thalamic volume as a potential primary endpoint for phase 2 proof-of-concept studies. These power calculations are different than those in our current literature for 2 reasons. First, whereas others have used discrete mathematical formulas,[40–42] we used a simulation-based approach. The simulation procedure takes advantage of our robust longitudinal dataset by using model parameters, variance-covariance matrix, and error variance that are estimated directly from a linear mixed effects model. The simulation procedure is based on an empirical distribution and was specifically designed to power calculate for longitudinal studies. Second, whereas others have assumed zero volume loss in the control group,[40–42] which may not be a biologically plausible assumption and could exaggerate study power (see Fig 4A), we incorporate data from healthy controls (ie, normal aging) into our calculations. Incorporating normal aging into sample size calculations is conceptually desirable, as the goal of a given neuroprotective drug should be to prevent MS-specific atrophy, with no expected impact of the drug on normal aging. As such, our power calculations are based on the concept of the “maximal effect size.” In the current era of DMTs, the maximal effect size may be attainable; recently presented data suggest that one of the currently available DMTs may reduce whole brain volume loss close to the range that is seen in normal aging beyond 2 years of therapy.[43] Our estimated sample size of 80 or 100 subjects per arm ($\alpha = 0.1$ or 0.05, respectively) in a 24-month study with 3 MRI time points is practical, and is similar to the sample size needed for phase 2 studies targeting new white matter lesions as a surrogate of preventing new attacks in phase 3 trials.[44] However, we recognize that, although robust, our dataset was from a single center; the variance in thalamic volume may increase when measured across multiple scanners, which would inflate the sample size needed to detect a given treatment effect. In addition, of necessity, the variance-covariance matrix used in our simulation procedure was estimated directly from

this dataset. Therefore, our sample size estimates pertain to the mix of subjects included in our dataset, and may differ when another mix of MS subjects is under consideration. As such, we provide sample size calculations as a preliminary finding that needs to be explored further in additional, multicenter datasets.

Our study has additional limitations. Our findings, particularly the consistency of thalamic volume decline across disease duration and subtypes, need to be replicated in other datasets. Our healthy controls had a shorter mean follow-up time than our MS cohort, which is reflected in the increased standard error around the estimate of thalamic volume decline in healthy controls. For future work, a better understanding of thalamic atrophy as part of normal aging from a larger sample of healthy controls with longer follow-up time will lead to a more accurate understanding of MS-specific thalamic atrophy and its potential utility as a clinical trial endpoint. A deeper understanding of MS-specific atrophy across the lifespan, including nonlinearity, may have implications for clinical trial design, as well as fundamental understanding of the disease. In addition, understanding the rate of thalamic atrophy in primary progressive MS would also be of interest. Finally, the incorporation of brain volume measurements into clinical practice is an interesting and complicated topic that is outside the scope of this work; several hurdles have been identified[45] and will need to be overcome to achieve this. Efforts have been made to develop practical methods to measure the thalamus specifically,[46] but further work is necessary to define the best segmentation and statistical methods before thalamic volume measurements can become a routine part of clinical care.

Thalamic atrophy could be used to substantially advance our understanding of progressive tissue loss in MS. Because of the potential implications of our findings, we felt it important to confirm their robustness with a second image-processing method that uses a different strategy to segment out the gray matter; we found excellent agreement between thalamic volume measurements derived from FIRST and FreeSurfer. This careful postprocessing analysis of standardized images at 3T, the large sample size of our study cohort, and the long disease duration represented in this analysis are strengths of this work. Finally, our power calculations used advanced methodology and a realistic assumption of an atrophy rate in the control group to address a practical and urgent need in the field. Accordingly, thalamic atrophy should remain under consideration as a potential biomarker reflecting neurodegeneration in MS and explored as an endpoint in drug trials aiming at neuroprotection.

Acknowledgment

This study received support from the NIH National Institute of Neurological Disorders and Stroke (R01NS062885 to D.P.) and the National Institute of Biomedical Imaging and Bioengineering (P41EB015922 to the University of Southern California Laboratory of Neuroimaging) for MRI data acquisition and processing, the National Multiple Sclerosis Society (Sylvia Lawry Physician Fellowship Award FP1778-A-1 to C.J.A.) and the Race to Erase MS Foundation (14–003399) to C.J.A.) for data analysis, and Biogen and GlaxoSmithKline for MRI data acquisition. This work was also supported by grant UL1TR001855 from the National Center for Advancing Translational Science (NCATS) of the NIH through an Institutional Career Development Award to C.J.A. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Trapp BD, Peterson J, Ransohoff RM, et al. Axonal transection in the lesions of multiple sclerosis. N Engl J Med 1998;338:278–285. [PubMed: 9445407]
2. Mahad DH, Trapp BD, Lassmann H. Pathological mechanisms in progressive multiple sclerosis. Lancet Neurol 2015;14:183–193. [PubMed: 25772897]
3. Haider L, Simeonidou C, Steinberger G, et al. Multiple sclerosis deep grey matter: the relation between demyelination, neurodegeneration, inflammation and iron. J Neurol Neurosurg Psychiatry 2014;85:1386–1395. [PubMed: 24899728]
4. Lassmann H. Multiple sclerosis: lessons from molecular neuropathology. Exp Neurol 2014;262(pt A):2–7. [PubMed: 24342027]
5. Witte ME, Mahad DJ, Lassmann H, van Horssen J. Mitochondrial dysfunction contributes to neurodegeneration in multiple sclerosis. Trends Mol Med 2014;20:179–187. [PubMed: 24369898]
6. Matute C, Domercq M, Sanchez-Gomez MV. Glutamate-mediated glial injury: mechanisms and clinical importance. Glia 2006;53: 212–224. [PubMed: 16206168]
7. Bjartmar C, Wujek JR, Trapp BD. Axonal loss in the pathology of MS: consequences for understanding the progressive phase of the disease. J Neurol Sci 2003;206:165–171. [PubMed: 12559505]
8. Cifelli A, Arridge M, Jezzard P, et al. Thalamic neurodegeneration in multiple sclerosis. Ann Neurol 2002;52:650–653. [PubMed: 12402265]
9. Vercellino M, Plano F, Votta B, et al. Grey matter pathology in multiple sclerosis. J Neuropathol Exp Neurol 2005;64:1101–1107. [PubMed: 16319720]
10. Wylezinska M, Cifelli A, Jezzard P, et al. Thalamic neurodegeneration in relapsing-remitting multiple sclerosis. Neurology 2003;60: 1949–1954. [PubMed: 12821738]
11. Henry RG, Shieh M, Okuda DT, et al. Regional grey matter atrophy in clinically isolated syndromes at presentation. J Neurol Neu-rosurg Psychiatry 2008;79:1236–1244.
12. Aubert-Broche B, Fonov V, Ghassemi R, et al. Regional brain atrophy in children with multiple sclerosis. Neuroimage 2011;58:409–415. [PubMed: 21414412]
13. Azevedo CJ, Overton E, Khadka S, et al. Early CNS neurodegeneration in radiologically isolated syndrome. Neuroimmunol Neuro-inflamm 2015;2:e102.
14. Henry RG, Shieh M, Amirbekian B, et al. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. J Neurol Sci 2009;282:61–66. [PubMed: 19394969]
15. Hammond KE, Metcalf M, Carvajal L, et al. Quantitative in vivo magnetic resonance imaging of multiple sclerosis at 7 tesla with sensitivity to iron. Ann Neurol 2008;64:707–713. [PubMed: 19107998]
16. Zivadinov R, Heininen-Brown M, Schirda CV, et al. Abnormal sub-cortical deep-gray matter susceptibility-weighted imaging filtered phase measurements in patients with multiple sclerosis: a case-control study. Neuroimage 2012;59:331–339. [PubMed: 21820063]
17. 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;21:1139–1150. [PubMed: 25583851]
18. 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;34:1931–1939. [PubMed: 23578679]
19. Houtchens MK, Benedict RH, Killiany R, et al. Thalamic atrophy and cognition in multiple sclerosis. Neurology 2007;69:1213–1223. [PubMed: 17875909]
20. Minagar A, Barnett MH, Benedict R, et al. The thalamus and multiple sclerosis: modern views on pathologic, imaging and clinical aspects. Neurology 2013;80:210–219. [PubMed: 23296131]
21. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald criteria”. Ann Neurol 2005;58:840–846. [PubMed: 16283615]
22. Mowry EM, Waubant E, McCulloch CE, et al. Vitamin D status predicts new brain magnetic resonance imaging activity in multiple sclerosis. Ann Neurol 2012;72:234–240. [PubMed: 22926855]

23. Lublin FD, Reingold SC, National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. The clinical course of multiple sclerosis: results of an international survey. Neurology 1996;46:907–911. [PubMed: 8780061]
24. Sdika M, Pelletier D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. Hum Brain Mapp 2009;30:1060–1067. [PubMed: 18412131]
25. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 2012;61:1402–1418. [PubMed: 22430496]
26. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 2011;56:907–922. [PubMed: 21352927]
27. McCulloch CE, Searle SR, Neuhaus JM. Generalized, linear, and mixed models. Hoboken, NJ: Wiley, 2008.
28. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J Educ Behav Stat 1998;24:323–355.
29. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33:1444–1452. [PubMed: 6685237]
30. Rudick RA, Antel J, Confavreux C, et al. Recommendations from the National Multiple Sclerosis Society Clinical Outcomes Assessment Task Force. Ann Neurol 1994;42:379–382.
31. Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust and automated longitudinal and cross-sectional brain change analysis. Neuroimage 2002;17:479–489. [PubMed: 12482100]
32. Psioda M Random Effects simulation for sample size calculations using SAS® In: Proceedings of the Southeast SAS Users Group. Cary, NC: SAS Institute, 2012.
33. Sharma MR, Karrison TG, Jim Y, et al. Resampling phase III data to assess phase II trial designs and endpoints. Clin Cancer Res 2012;18:2309–2315. [PubMed: 22287601]
34. Hedman AM, van Haren NE, Schnack HG, et al. Human brain changes across the life span: a review of 56 longitudinal magnetic resonance imaging studies. Hum Brain Mapp 2012;33: 1987–2002. [PubMed: 21915942]
35. Vollmer T, Signorovitch J, Huynh L, et al. The natural history of brain volume loss among patients with multiple sclerosis: a systematic literature review and meta-analysis. J Neurol Sci 2015; 357:8–18. [PubMed: 26238166]
36. Kolasinski J, Stagg CJ, Chance SA, et al. A combined postmortem magnetic resonance imaging and quantitative histological study of multiple sclerosis pathology. Brain 2012;135:2938–2951. [PubMed: 23065787]
37. Benedict RH, Ramasamy D, Munschauer F, et al. Memory impairment in multiple sclerosis: correlation with deep grey matter and mesial temporal atrophy. J Neurol Neurosurg Psychiatry 2009;80: 201–206. [PubMed: 18829629]
38. Batista S, Zivadinov R, Hoogs M, et al. Basal ganglia, thalamus and neocortical atrophy predicting slowed cognitive processing in multiple sclerosis. J Neurol 2012;259:139–146. [PubMed: 21720932]
39. Sprenger T, Gaetano L, Radue EW, et al. Fingolimod reduces deep grey matter and regional volume loss in the brain of RRMS patients: a post-hoc analysis of FREEDOMS and FREEDOMS II data. ECTRIMS Online Library 2015;115986.
40. Altmann DR, Jasperse B, Barkhof F, et al. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. Neurology 2009;72:595–601. [PubMed: 19005170]
41. Van den Elskamp IJ, Boden B, Dattola V, et al. Cerebral atrophy as outcome measure in short-term phase 2 clinical trials in multiple sclerosis. Neuroradiology 2010;52:875–881. [PubMed: 20049424]
42. Jones BC, Nair G, Shea CD, et al. Quantification of multiple-sclerosis-related brain atrophy in two heterogeneous MRI datasets using mixed-effects modeling. Neuroimage 2013;3:171–179. [PubMed: 24179861]
43. Barkhof F, Cohen JA, Coles AJ, et al. on behalfof the CARE-MS I and CARE-MS II Investigators. Alemtuzumab slows brain volume loss over 5 years in patients with active relapsing-remitting multiple sclerosis with most patients not receiving treatment for 4 years: CARE MS I and II extension study. ECTRIMS Online Library 2015;116687.

44. Hauser SL, Waubant E, Arnold DL, et al. B-cell depletion with rit-uximab in relapsing-remitting multiple sclerosis. N Engl J Med 2008;358:676–688. [PubMed: 18272891]

45. Azevedo CJ, Pelletier D. Whole brain atrophy: ready for implementation into clinical decision-making in multiple sclerosis? Curr Opin Neurol 2016;29:237–242. [PubMed: 27070219]

46. Menendez-Gonzalez M, Salas-Pacheco JM, Arias-Carrion O. The yearly rate of relative thalamic atrophy (yrRTA): a simple 2D/3D method for estimating deep gray matter atrophy in multiple sclerosis. Front Aging Neurosci 2014;6:219. [PubMed: 25206331]

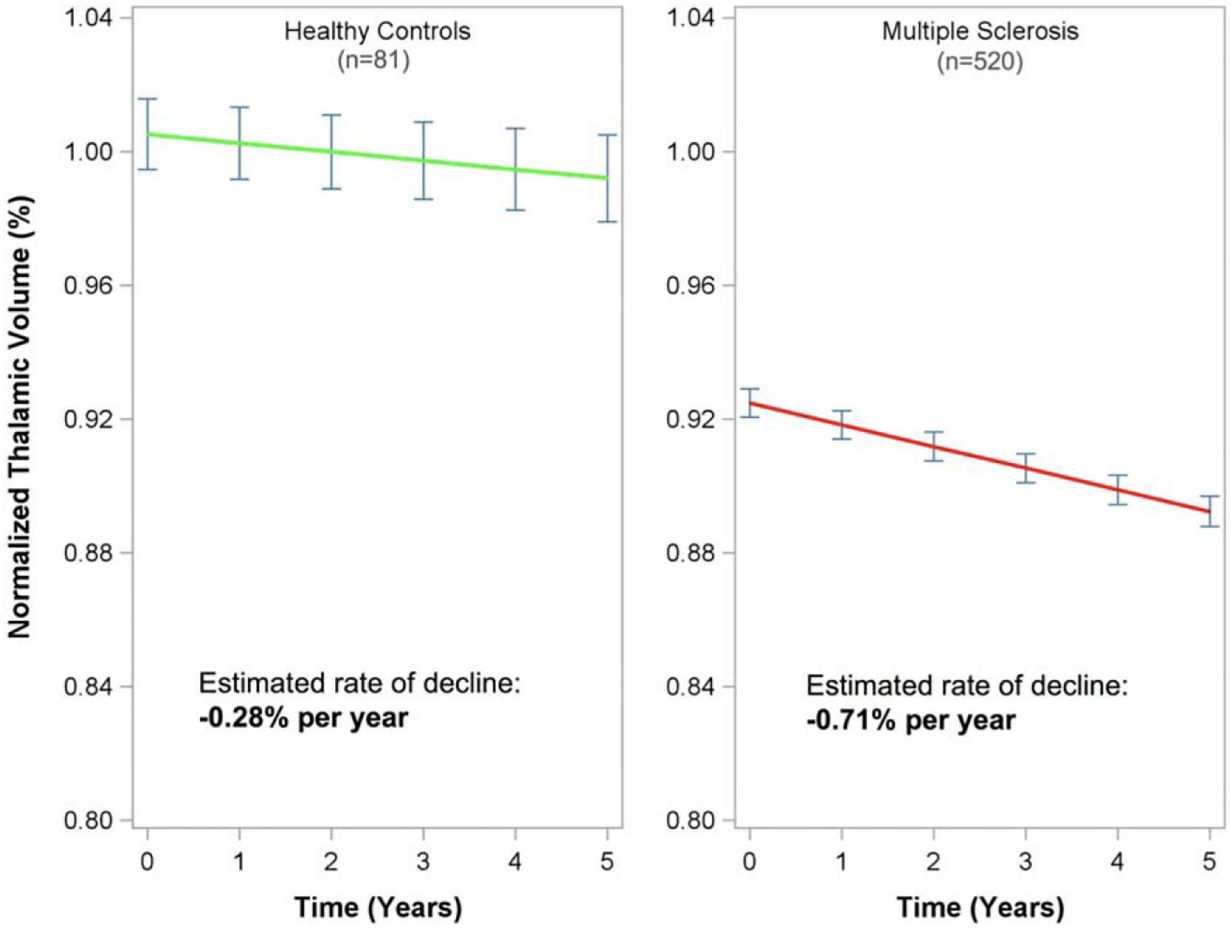

FIGURE 1:

Normalized thalamic volume decline from a linear mixed effects model including both healthy controls (green line, left panel) and multiple sclerosis subjects (red line, right panel) is shown. Point estimates of normalized thalamic volume at each year ±95% standard error (blue error bars) are shown.

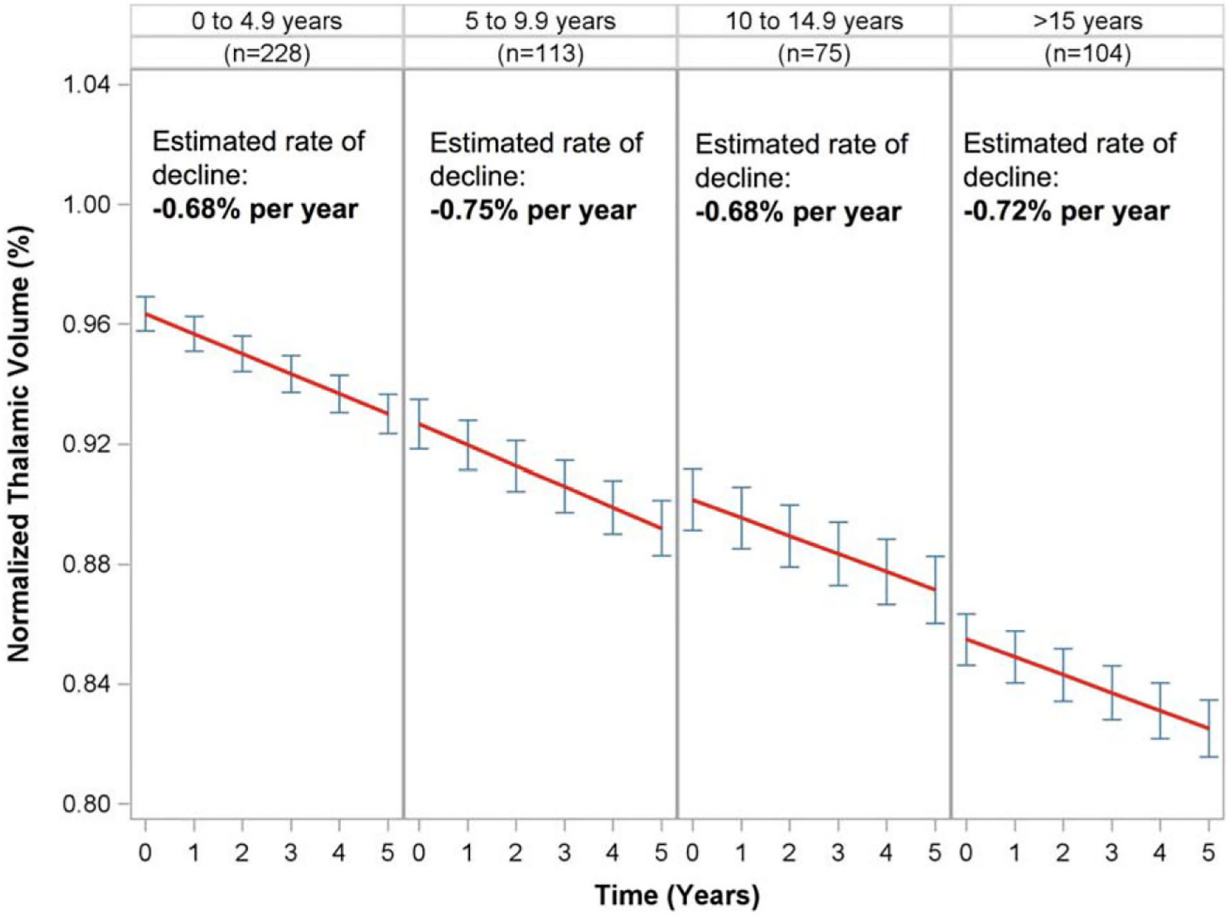

FIGURE 2:

Normalized thalamic volume over time in multiple sclerosis subjects divided into 5-year epochs of disease duration at study entry. Estimated decline in each epoch is shown (red lines), with point estimates of normalized thalamic volume±95% standard error at each year (blue error bars).

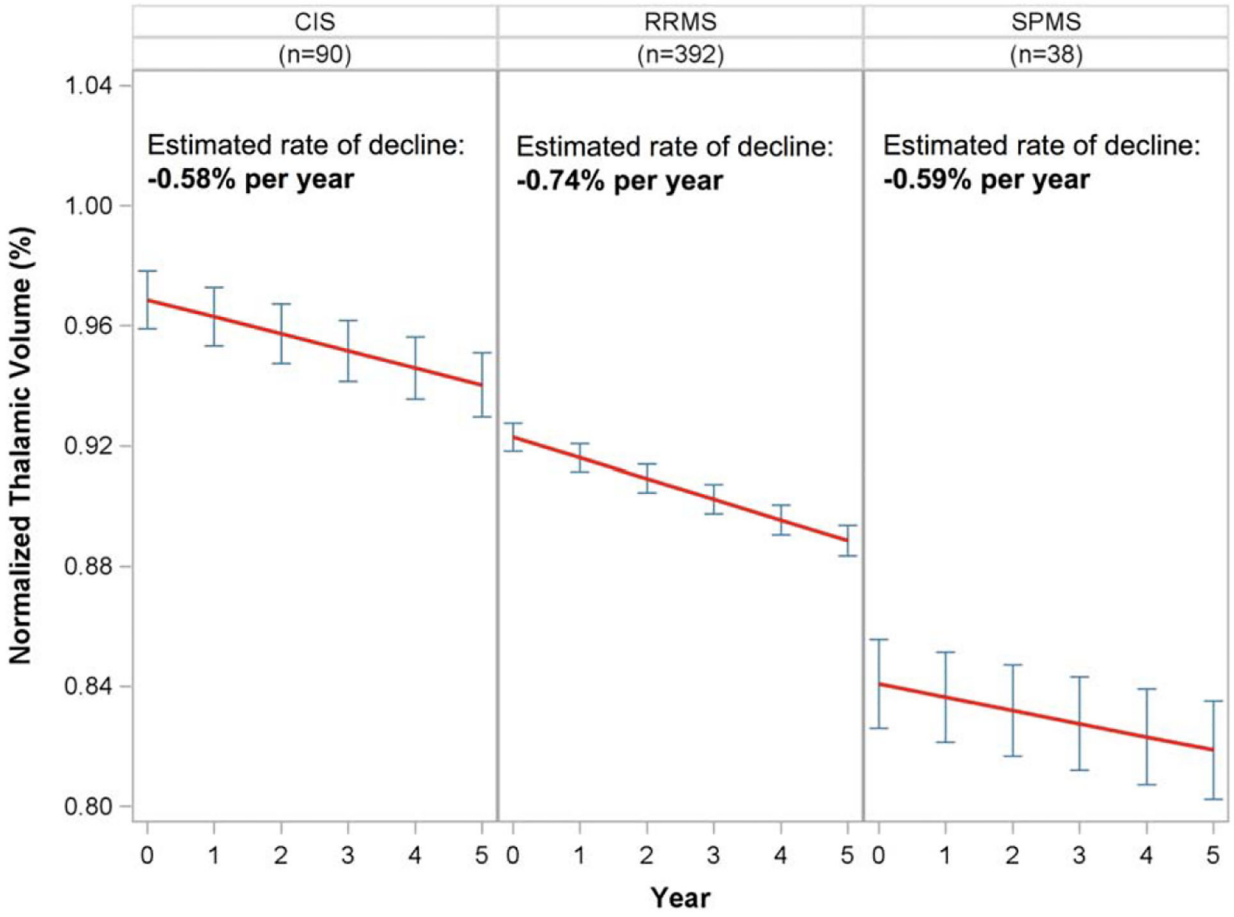

FIGURE 3:

Normalized thalamic volume by clinical subtypes at baseline. Estimated decline in each subtype from linear mixed effects models is shown (red lines), with point estimates of normalized thalamic volume±95% standard error at each year (blue error bars). CIS=clinically isolated syndrome; SPMS=secondary progressive multiple sclerosis; RRMS=relapsing–remitting multiple sclerosis.

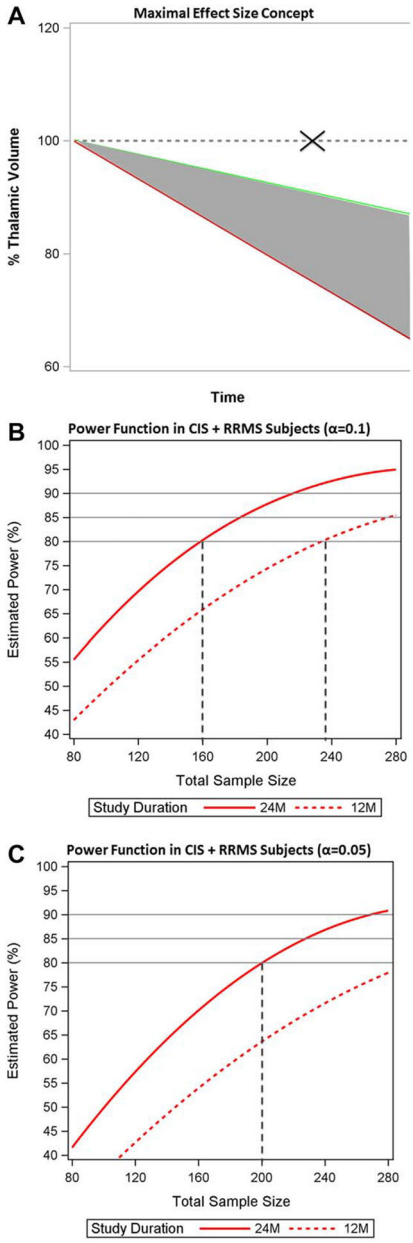

FIGURE 4:

Maximal effect size concept and sample size estimations. (A) The concept of maximal effect size. The rate of thalamic volume loss, estimated from linear mixed effects models, is shown in healthy controls (green line) and multiple sclerosis (MS) subjects (red line). “Maximal effect size” refers to a hypothetical treatment that could slow the rate of thalamic atrophy in MS to that of normal aging, that is to say the MS-specific atrophy that is amenable to intervention. The dashed line represents the trajectory that would occur if there were zero atrophy in the control group, an assumption not supported by the data (marked with an X). (B) The power function from a simulation procedure estimating the total sample size that would be needed to detect the maximal effect size in clinically isolated syndrome (CIS) and relapsing–remitting MS (RRMS) subjects, assuming α=0.1. To achieve 80% power, 160

subjects total would be needed in a 24-month study (80 per arm, solid red curve) and 235 subjects total in a 12-month study (118 per arm, dashed red curve). (C) The power function from a simulation procedure estimating the total sample size that would be needed to detect the maximal effect size in CIS and RRMS subjects, assuming α=0.05. To achieve 80% power, 200 subjects total would be needed in a 24-month study (100 per arm, solid red curve) and >280 subjects total in a 12-month study (dashed red curve).

TABLE 1.

Subject Demographics and Characteristics

Variable	MS Subjects, n = 520[a]	Healthy Controls, n = 81[a]	p[b]
Age at study entry, yr	42.7 ± 9.8	41.1 ± 9.7	0.2
Gender			0.4
Female	365 (70.2)	53 (65.4)	
Male	155 (29.8)	28 (35.6)	
Disease course		N/A	
CIS	90 (17.3)		
RRMS	392 (75.4)		
SPMS	38 (7.3)		
Disease duration, yr	9.2 ± 8.6	N/A	
Mean EDSS at baseline	1.8 ± 1.5	N/A	
Cumulative DMT exposure		N/A	
None	116 (22.3)		
≤1 year	33 (6.3)		
1–3 years	77 (14.8)		
3–5 years	107 (20.6)		
>5 years	187 (36.0)		
Mean follow-up time, yr	4.1 ± 1.6	1.3 ± 1.7	
Total MRI time points analyzed	2,485	147	
MRI time points, No.			
Subjects with 1 MRI	18	41	
Subjects with 2 MRIs	46	19	
Subjects with 3 MRIs	51	17	
Subjects with 4 MRIs	49	3	
Subjects with 5 MRIs	110	1	
Subjects with 6 MRIs	246	—	

[a]Mean±standard deviation for continuous variables, No. (%) for categorical variables.

Ann Neurol. Author manuscript; available in PMC 2019 January 03.

[b] *p*-value for t test or chi-square test.

CIS=clinically isolated syndrome; DMT=disease-modifying therapy; EDSS=Expanded Disability Status Scale; MRI=magnetic resonance imaging; MS=multiple sclerosis; N/A=not applicable; RRMS=relapsing–remitting multiple sclerosis; SPMS=secondary progressive multiple sclerosis.

TABLE 2.

Correlation between Thalamic Volume and Clinical Endpoints

	EDSS	MSFC z Score	PASAT	9-Hole Peg Test	T25FW
Thalamic volume	−0.29	0.32	0.15	−0.37	−0.25
	(−0.21 to −0.37)	(0.24 to 0.40)	(0.06 to 0.23)	(−0.30 to −0.45)	(−0.16 to −0.32)
	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$

Association between thalamic volume and clinical endpoints at baseline. Values presented are Spearman correlation coefficent (95% confidence interval) and p-value.

EDSS=Expanded Disability Status Scale; MSFC=Multiple Sclerosis Functional Composite; PASAT=Paced Auditory Serial Addition Test; T25FW=Timed 25-Foot Walk.

TABLE 3.

Correlation between Percentage Change in Thalamic Volume and Percentage Change in Whole Brain Volume

	% Change in Whole Brain Volume over 2 Years	**% Change in Whole Brain Volume over 5 Years**
% Change in thalamic volume over 1 year	0.35 (0.26–0.44)	0.23 (0.12–0.34)
	$p < 0.0001$	$p < 0.0001$
% Change in thalamic volume over 2 years	0.60 (0.52–0.67)	0.40 (0.29–0.51)
	$p < 0.0001$	$p < 0.0001$

The association between percentage change in thalamic volume and percentage change in whole brain volume measured with Spearman correlation. Values presented are point estimate (95% confidence interval) and p-value.