Published in final edited form as:
J Neurol Neurosurg Psychiatry. 2009 November ; 80(11): 1218–1223. doi:10.1136/jnnp.2008.154732.

Sample size requirements for treatment effects using gray matter, white matter and whole brain volume in relapsing–remitting multiple sclerosis

B Healy[1,2], **P Valsasina**[3], **M Filippi**[3], and **R Bakshi**[4]

[1] Department of Neurology, Brigham and Women's Hospital, Partners MS Center, Harvard Medical School, Boston, Massachusetts, USA

[2] Department of Neurology, Biostatistics Center, Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA

[3] Neuroimaging Research Unit, Scientific Institute and University Ospedale San Raffaele, Milan, Italy

[4] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Abstract

Objective—To compare the sample size requirements for a neuroprotection trial with change in cerebral gray matter volume (GMV), white matter volume (WMV) or whole brain parenchymal volume (BPV) as outcome measures in patients with relapsing–remitting multiple sclerosis (RRMS).

Methods—Two datasets with longitudinal MRI measures of untreated patients with RRMS (n = 116 and n = 26) and one dataset of treated patients with RRMS (n = 109) were investigated. In each dataset, normalised GMV, normalised WMV and normalised BPV were analysed using a random intercepts and slopes model to estimate the variance components and per cent change. The required sample size to observe a 33%, 50% and 90% reduction in the per cent change was calculated for each dataset using both a constant per cent change for each measurement and the estimated per cent change for each dataset.

Results—The per cent change was greatest in GMV but all variance components were smallest in BPV. Using the estimated per cent change, the sample size required in the untreated cohorts was similar for GMV and BPV, and both were lower than WMV. In the treated cohort, the sample size for GMV was the smallest of all measures. Including additional scans reduced the sample size but increasing the length of the trial and clustering scans led to greater reductions.

Conclusions—Cerebral GMV may be a viable outcome measure for clinical trials investigating neuroprotection in RRMS patients, especially considering that the treatment effect may be larger on GMV compared with BPV. However, GMV was somewhat limited by increased variability versus BPV.

Correspondence to: Dr R Bakshi, Brigham and Women's Hospital, Laboratory for Neuroimaging Research, 1, Brookline Place, Suite 602, Brookline, MA 02445, USA; rbakshi@bwh.harvard.edu.

Supplementary data are published online only at http://jnnp.bmj.com/content/volXX/issueX

Competing interests: None.

Ethics approval: Ethics approval was obtained.

The primary end point of most clinical trials in relapsing–remitting multiple sclerosis (RRMS) has been a marker of inflammation. For phase I/II clinical trials, this end point is usually gadolinium enhancing lesions on brain MRI scans; in phase III trials, the number of clinical relapses or development of disability is the most common primary end point, with MRI measures providing secondary and tertiary end points. Although FDA approved treatments have shown efficacy in reducing inflammation, the treatments have limited effects in reducing the destructive aspects of the disease, as assessed, for example, by brain or spinal cord atrophy. One explanation for this is that MS consists of two processes—an inflammatory process and a degenerative process—and the present treatments target the inflammatory process. Although the treatments lead to a reduction in relapse rate, the degenerative phase of the disease may not be fully interrupted. This problem is prevalent for patients with both the RR and progressive forms of MS.

Research now focuses on a new class of treatments that target neuroprotection, rather than reduction of inflammation.[1–4] All MS patients are likely to benefit from effective neuroprotective treatments because CNS atrophy begins early in the disease process.[5] As neuroprotection cannot be measured clinically over the length of time for most clinical trials, one attractive surrogate for measuring such efficacy is by MRI based measures of brain and spinal cord atrophy. The cerebral neurodegeneration measurements of interest include whole brain atrophy, measured by reduction in normalised brain parenchymal volume (BPV), or compartmental tissue atrophy, measured by reduction in normalised gray matter volume (GMV) or white matter volume (WMV). Treatments aimed at mitigating brain atrophy may be effective for MS, particularly when combined with immunomodulatory treatments.

A growing body of evidence indicates gray matter involvement in RRMS such as lesions, atrophy, abnormal neuronal metabolites, reduced magnetisation transfer ratio, increased diffusivity and T2 hypointensity.[6] Several cross sectional studies have linked gray matter MRI involvement with clinical impairment, including physical disability and cognitive dysfunction.[7–11] These findings underscore the potential for gray matter degeneration to serve as a marker of disease progression in MS. As recently reviewed, gray matter disproportionately undergoes atrophy compared with white matter in the MS brain.[9 12–14] Moreover, WMV is susceptible to fluctuations based on inflammatory activity such as increased volume that may offset degenerative effects.[13 15] Thus GMV loss may represent a more sensitive tool for assessing neuroprotective therapeutic effects than BPV or WMV measures.

In order to determine the potential efficacy of neuroprotective agents, clinical trials need to be designed, and the accompanying sample size calculations are required. Sample size calculations for the effect of neuroprotective agents on whole brain atrophy have recently been described in a study demonstrating the reduction in sample size via registration based MRI segmentation.[16] Because of the limited availability and technical challenges associated with registration based serial analysis of GMV and WMV, we chose to compare the sample size requirements for gray, white and whole brain volume based on serial measurements of normalised volumes. We assessed both a monotherapy versus untreated study design and a combination trial in which platform therapy is combined with a second therapy targeted to reduce brain atrophy. The calculations in this paper are based on an exponential decline in brain volume over the course of observation but a linear decline is a simple extension of the method presented.

Materials and Methods

Subjects

Three datasets with longitudinal MRI measures of RRMS patients were investigated; two of untreated patients and one of treated patients. The first untreated dataset ($RRMS_1$) was a

placebo group from a clinical trial scanned monthly for 9 months (n = 116), and the second untreated dataset ($RRMS_2$) was an observational cohort scanned every 3 months for 1.5 years (n = 26).[14 17] The treated dataset ($RRMS_3$) was the group receiving daily glatiramer acetate for 9 months (n = 109) from the same clinical trial as $RRMS_1$.[14] While we have data on treatment versus placebo in this comparison trial, the purpose of the present study was to use these data only to determine sample sizes for future studies and not to evaluate treatment efficacy on GMV or WMV from the trial. We plan to present such an analysis in a separate publication. The enrolment criteria for each of these datasets have been previously published, and the demographics of the patients in each dataset are provided in table 1. Although each group was classified as RRMS, the second untreated cohort consisted of patients with longer disease duration and lower expanded disability status score.[18] In this cohort, patients voluntarily refused therapy for the duration of the study; therefore, these patients may have been more benign than the general RRMS population.

MRI acquisition and segmentation

All datasets consisted of two-dimensional T1 weighted spin echo images acquired at 1.5 T. All images had a field of view of 250 mm, matrix of 256×256 and pixel size = 0.98×0.98 mm^2. The $RRMS_1$ and $RRMS_3$ datasets consisted of 44 contiguous axial slices covering the entire brain with a slice thickness of 3 mm. The images were part of a multicentre study in which the allowed ranges of contrast parameters were the following: TR = 450–650 ms, TE = 10–20 ms. The $RRMS_2$ dataset consisted of 24 contiguous axial slices covering the entire brain with a slice thickness of 5 mm. The images were acquired in San Raffaele hospital using a TR of 768 ms and a TE of 15 ms.

SIENAX was used to estimate the three brain volume measurements: normalised GMV, normalised WMV and normalised BPV.[19] SIENAX uses a brain extraction tool to segment brain from non-brain tissue in the head and to estimate the skull surface. Then, the extracted image was segmented into WM, GM and CSF, to estimate the absolute volumes of the tissue compartments. The original image was registered to a canonical image in a standardised space (derived from the MNI152 standard space) by using both the brain image and the exterior skull surface, to provide a spatial normalisation scaling factor for each patient. The estimated absolute volumes were then multiplied by the normalisation factor to yield normalised parenchymal volumes of these tissue compartments. The output results from all segmentations were visually inspected, and there were no severe effects of misclassifications.

Statistical methods

The demographic and baseline characteristics of the $RRMS_1$ and $RRMS_2$ cohorts were compared using a Wilcoxon rank sum test; $RRMS_3$ was not significantly different from $RRMS_1$ in any characteristic because these datasets were two arms from the same clinical trial. Brain volumes were also compared, controlling for disease duration using linear regression. A mixed effects model with a random intercept and slope was fit to the longitudinal brain matter volume data (equation 1).[20]

$$\log(Vol_{i,j})=\beta_0+\beta_1 year+b_{0,i}+b_{1,i}year+e_{i,j} \quad \text{Equation 1}$$

Since the focus of the analysis was per cent change, the mixed effects model was fit on the log scale. A transformation of the regression parameter, β_1, was used to find the average per cent change. A treatment that reduces the per cent change over time would lead to a difference in the slope of the regression line from equation 1. The SD of the random intercept (σ_{ri}) and slope (σ_{rs}), the correlation between the random intercept and slope (ρ_r) and the residual SD (σ_e) were

estimated from this model and utilised in the power calculation. As outliers can have a large effect on regression models, severe outlying values were removed from the analysis. A severe outlying value was defined a priori as a change of more than 0.1 on the log scale on consecutive scans for either GMV or WMV or a change of more than 0.05 on the log scale on consecutive scans for the BPV because these differences were likely due to mechanical error rather than true change. For $RRMS_1$, one patient was fully removed, and an additional 32 scans on 24 patients were removed; for $RRMS_2$, one patient was fully removed, and an additional eight scans on seven patients were removed. Finally, for $RRMS_3$, three patients were fully removed, and an additional 27 scans on 21 patients were removed. The cause of the outlying scans in all cases was technical in nature, due to either inhomogeneity between the two interleaved sets of images or poor segmentation of infratentorial regions.

A program for sample size calculation for differences in slopes with longitudinal data is freely available online (http://hedwig.mgh.harvard.edu/biostatistics/software). Sample size calculations for GMV, WMV and BPV as candidate outcome measures were based on the estimate of the per cent change from each individual cohort and on an external value derived from a recent literature review (0.7% annual decrease for all three volume measurements).[12] In all cases, we estimated the sample size required to detect a 33%, 50% and 90% reduction in the per cent change with 80% power at the two sided 0.05 level using a study with MRI measurements at baseline, year 1 and year 2, and no dropouts. The reported sample sizes were the total numbers of patients required assuming a 1:1 allocation of patients to treatment and placebo arms. The effect of changing the number of scans and changing the total length of follow-up were investigated. Clustering of scans was also investigated based on a previously proposed paradigm.[21]

Results

Relationship between MRI and disease type

Baseline GMV, WMV and BPV were compared across groups (table 1). Although WMV was significantly different in univariate analysis, the difference did not remain after adjusting for disease duration ($p = 0.21$). All univariate and adjusted comparisons of GMV and BPV were not significant.

Model characteristics

The estimated per cent change and variance components from the model in equation (1) for each patient population were calculated (table 2). The per cent change in all groups was highest in GMV, and this change was much greater than the 0.7% decrease from external data. WMV changed less than GMV in the $RRMS_2$ dataset and actually increased in the $RRMS_1$ and $RRMS_3$ groups. The change in BPV was always between the change in WMV and GMV. Finally, all variance components were lowest for BPV in each cohort.

Sample size calculation

The sample size requirements based on the estimated per cent change are provided in table 3. In the $RRMS_1$ cohort, the ratio of the sample size for GMV and BPV was 1:1.8 while in the $RRMS_2$ cohort, the ratio of the sample size for GMV and BPV was 1.6:1. In each cohort, the sample size requirements for WMV were much greater than for either GMV or BPV. In the $RRMS_3$ cohort, BPV required more than a fivefold increase in sample size compared with GMV but the sample size requirements for GMV were similar to $RRMS_1$. As WMV increased over the course of the study in the $RRMS_1$ and $RRMS_3$ cohorts, the reduction in the per cent change corresponded to the treatment actually leading to a smaller increase in WMV, rather than a smaller decrease as for the other measures. Although this treatment effect is not of interest, a similar sample size is required if the magnitude of the per cent decrease in WMV

was the same as the estimated per cent increase. The sample size calculations based on 0.7% decrease for each measurement are available in the online supplement accompanying this paper. In all cohorts, BPV required the smallest sample sizes because this measurement had the smallest variance.

Increasing the number of on-study MRI scans during the trial had a limited effect on the required sample size but increasing the length of the trial by 1 year decreased the required sample size by approximately 10% (table 4). As previously demonstrated,[21] designs that clustered observations at the beginning and end of the study required lower sample sizes compared with designs with equally spaced scans. The effects of the number of scans and the length of the trial were similar for the different brain volume measurements.

Discussion

The results in this paper show the viability of compartment based global brain volume measurements as outcomes in clinical trials for neuroprotection in RRMS patients. Although we report the sample size calculations based on both the estimated per cent change (table 3) and a constant per cent change from previous work (see supplementary data available online), the calculations based on the estimated per cent change were likely more accurate because the variance components and the per cent change were estimated from the same data. As the estimated per cent change was larger in each cohort than the assumed constant per cent change, the variance was likely also higher in our cohorts compared with previous cohorts. In addition, the assumption of an equal per cent change of 0.7% for each compartment may misrepresent the true level of gray and white matter atrophy. In one study, the rate of volume change was similar in gray versus white matter[22] but the gray matter volume change was observed to be higher than the white matter volume change in two other studies.[23,24] Therefore, the results from table 3 may be more appropriate for design of future trials.

Although BPV had reduced variance because misclassification of CSF is less common than misclassification of either GMV or WMV,[9] combining the two compartments of brain tissue reduced the information available if the disease was disproportionately affecting GMV or WMV. For example, if MS disproportionately caused GMV loss, the change in BPV would be contaminated by WMV, which is prone to fluctuations and thus relatively insensitive and non-specific for axonal loss. It has been postulated that increases in WMV due to axonal swelling, inflammation or oedema may offset any disease related WMV atrophy.[6,13,15] It should be kept in mind that it has been hypothesised but not unequivocally proven that GMV is a more stable and reliable measure of neuroprotection and axonal loss than WMV. It remains possible that GMV may also be affected by transient fluctuations separate from tissue destructive processes.

The differences in the required sample size for the two untreated cohorts may be related to the heterogeneity of the patient populations. $RRMS_1$ was a placebo group preselected for a clinical trial with relatively early and active MS. Conversely, $RRMS_2$ was an observational untreated group with a later stage of MS. $RRMS_1$ is likely more representative of those who would be enrolled in a neuroprotection clinical trial; therefore, the results for this cohort are likely most relevant for design of future trials. At the same time, $RRMS_2$ provides information for design of trials in patients with more benign MS. Even though these patients have a less severe disease course, they still have brain atrophy and may benefit from neuroprotective therapy.

Comparisons of the treated ($RRMS_3$) versus the untreated ($RRMS_1$) patients from the same population showed that the sample sizes for GMV were similar but the sample size for BPV increased almost fourfold. The increase in sample size for BPV was due to the difference in estimated per cent change (table 2). A 50% reduction in the per cent change is smaller in

magnitude in the treated group compared with the untreated group; therefore, a larger sample size is required to observe the effect. The smaller increase in sample size for GMV was due to the more modest difference in the estimated per cent change comparing the treated and untreated groups (table 2) such that a 50% reduction in the per cent change was similar in magnitude in both groups. The estimated variability in each measurement was similar for each group. Therefore, GMV may be an even more promising marker for neuroprotection trials if a combination strategy is employed where both treatment arms are receiving platform immunomodulatory therapy.

Up to now, we assumed that the treatment would lead to the same reduction in the per cent change for each of the brain volume measurements. A recent study has suggested that the treatment effect of interferon β1a on preserving GMV is much greater than the effect on either BPV or WMV.[22] As the treatment effect may be greater on GMV, the sample size requirement for a clinical trial involving GMV may correspond to a larger reduction in per cent change than in either BPV or WMV. For example, a treatment may lead to a 50% reduction in the per cent change in GMV while leading to only a 33% reduction in the per cent change in BPV. In this case, GMV would require a smaller sample size in all RRMS cohorts when the estimated per cent change is considered.

The large decrease in GMV contrasted with either increases or only slight decreases in WMV were consistent with other cohorts of clinically isolated syndrome (CIS) or RRMS patients.[13,23,24] As the behaviour of the cohorts was similar to some previous RRMS cohorts, our results can likely be generalised to other RRMS populations. However, in examining the results from all longitudinal GMV and WMV studies, there is variability in atrophy rates across studies[12]; thus we urge caution until further studies are available to confirm and extend our results. For example, one study has shown similar change in the gray and white matter volume in untreated patients.[22] The increase in the size of the white matter over the course of the study in the $RRMS_1$ and $RRMS_3$ cohort is consistent with observations made in CIS patients, which is likely related to the effect of inflammation/plaques and oedema increasing WMV.[13] The increase in WMV may also be related to error in measurement such as misclassification of gray matter as white matter. Therefore, the estimated per cent decrease in GMV from this sample may be overestimating the true level.

Beyond misclassification, several observations were considered outliers based on large changes in one measurement. In all cases, the source of error was associated with imaging related technical factors. The errors associated with inhomogeneities in the image intensity might be related to the coil used for acquisition or to other scanner related noise sources. If one wants to use GMV and WMV as outcomes in clinical trials, one should check carefully that all centres apply image intensity correction and include a sequence optimised for gray–white segmentation (eg, MP-RAGE). Our findings should be interpreted in the context of previous observations indicating that technical factors such as imaging segmentation methodology play a major role in the results obtained in assessing atrophy rate and treatment effects.[25] For example, it is conceivable that other methods of performing cerebral gray versus white matter image segmentation, such as FreeSurfer,[26] with appropriate source images, would yield different sample sizes.

When the effects of performing more frequent scanning during the 2 year observation period were investigated, the reductions in sample size were modest. At the same time, clustering measurements at the beginning and end of the trial led to a decrease in the sample size requirements. The optimal spacing and number of scans must be determined for each trial based on the length of follow-up, feasibility, expected dropout rate and within versus between person variability. An excellent discussion of the different forms of variability in atrophy measures and the benefits and difficulties related to clustering scans at the beginning and end of a trial

has recently been published for whole brain atrophy measures, but the discussion applies equally to our paradigm.[21]

The sample size calculations from our results are similar to those derived from the segmented brain volume difference observed by another group.[16] The significant reductions in sample size observed from the registration techniques (eg, SIENA[19]) are not available for gray matter and white matter because registration techniques are only widely used for whole brain atrophy. If registration techniques were improved for GMV and WMV such that these techniques could be applied on a large scale, similar reductions in sample size would be expected for these two measures. At the same time, the sample size requirements observed in table 3 demonstrate the viability of these measurements in clinical trials using the present technology.

An important limitation of our study is that our sample size calculations only apply to RRMS patients who are untreated or starting glatiramer acetate, but all MS patients will be candidates for these types of therapies. We urge caution in attempting to extend our results to other MS phenotype groups such as patients with primary or secondary progressive MS, due to the likely differences in the rate of compartmental atrophy versus patients with RRMS and CIS.[12] It will also be of interest to compare our findings with patients receiving other disease modifying therapies such as interferon-β. An additional limitation of our study is the short duration of treatment data (9 months) compared with the length of trial proposed (2 years). Our model assumes the atrophy rate is equal over the course of treatment but immunomodulatory treatment may not reduce atrophy immediately after initiation. Rather, the treatment may require several months before an effect is observed.[27,28] As patients in our study were starting immunomodulatory treatment, the observed per cent change in our study may be higher than if further follow-up data had been investigated. In addition, if patients were observed once the treatment was in full effect, the change over time may have been reduced further. Due to these factors, our sample size estimates for treated patients may be underestimated, but there is no evidence that the relative difference between the compartmental atrophy and the whole brain atrophy would be affected.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: BH acknowledges support from the Partners MS Center. RB acknowledges support of research grants from the National Institutes of Health (1R01NS055083-01) and National Multiple Sclerosis Society (RG3705A1; RG3798A2). Teva kindly accepted to use the placebo and treated data of the European/Canadian MRI trial of Copaxone in MS.

References

1. Waxman SG. Mechanisms of disease: sodium channels and neuroprotection in multiple sclerosis-current status. Nat Clin Pract Neurol 2008;4:159–69. [PubMed: 18227822]
2. Dhib-Jalbut S, Arnold DL, Cleveland DW, et al. Neurodegeneration and neuroprotection in multiple sclerosis and other neurodegenerative diseases. J Neuroimmunol 2006;176:198–215. [PubMed: 16983747]
3. Kalkers NF, Barkhof F, Bergers E, et al. The effect of the neuroprotective agent riluzole on MRI parameters in primary progressive multiple sclerosis: a pilot study. Mult Scler 2002;8:532–3. [PubMed: 12474997]
4. Stankiewicz J, Panter SS, Neema M, et al. Iron in chronic brain disorders: imaging and neurotherapeutic implications. Neurotherapeutics 2007;4:371–86. [PubMed: 17599703]

5. Filippi M, Rovaris M, Inglese M, et al. Interferon beta-1a for brain tissue loss in patients at presentation with syndromes suggestive of multiple sclerosis: a randomised, double-blind, placebo-controlled trial. Lancet 2004;364:1489–96. [PubMed: 15500893]
6. Pirko I, Lucchinetti CF, Sriram S, et al. Gray matter involvement in multiple sclerosis. Neurology 2007;68:634–42. [PubMed: 17325269]
7. Brass SD, Benedict RH, Weinstock-Guttman B, et al. Cognitive impairment is associated with subcortical magnetic resonance imaging grey matter T2 hypointensity in multiple sclerosis. Mult Scler 2006;12:437–44. [PubMed: 16900757]
8. Zhang Y, Zabad RK, Wei X, et al. Deep gray matter “black T2” on 3 tesla magnetic resonance image correlates with disability in multiple sclerosis. Mult Scler 2007;13:880–3. [PubMed: 17468444]
9. 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. [PubMed: 15961046]
10. Amato MP, Bartolozzi ML, Zipoli V, et al. Neocortical volume decrease in relapsing–remitting MS patients with mild cognitive impairment. Neurology 2004;63:89–93. [PubMed: 15249616]
11. Sanfilipo MP, Benedict RH, Weinstock-Guttman B, et al. Gray and white matter brain atrophy and neuropsychological impairment in multiple sclerosis. Neurology 2006;66:685–92. [PubMed: 16534104]
12. Bermel RA, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. Lancet Neurol 2006;5:158–70. [PubMed: 16426992]
13. Dalton CM, Chard DT, Davies GR, et al. Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. Brain 2004;127:1101–7. [PubMed: 14998914]
14. Valsasina P, Benedetti B, Rovaris M, et al. Evidence for progressive gray matter loss in patients with relapsing–remitting MS. Neurology 2005;65:1126–8. [PubMed: 16217074]
15. Fisher E, Chang A, Fox RJ, et al. Imaging correlates of axonal swelling in chronic multiple sclerosis brains. Ann Neurol 2007;62:219–28. [PubMed: 17427920]
16. Anderson VM, Bartlett JW, Fox NC, et al. Detecting treatment effects on brain atrophy in relapsing remitting multiple sclerosis: Sample size estimates. J Neurol 2007;254:1588–94. [PubMed: 17940723]
17. Oreja-Guevara C, Rovaris M, Iannucci G, et al. Progressive gray matter damage in patients with relapsing–remitting multiple sclerosis: a longitudinal diffusion tensor magnetic resonance imaging study. Arch Neurol 2005;62:578–84. [PubMed: 15824256]
18. Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33:1444–52. [PubMed: 6685237]
19. Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 2002;17:479–89. [PubMed: 12482100]
20. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982;38:963–74. [PubMed: 7168798]
21. Schott JM, Frost C, Whitwell JL, et al. Combining short interval MRI in Alzheimer's disease: implications for therapeutic trials. J Neurol 2006;253:1147–53. [PubMed: 16998650]
22. Zivadinov R, Locatelli L, Cookfair D, et al. Interferon beta-1a slows progression of brain atrophy in relapsing–remitting multiple sclerosis predominantly by reducing gray matter atrophy. Mult Scler 2007;13:490–501. [PubMed: 17463072]
23. Tiberio M, Chard DT, Altmann DR, et al. Gray and white matter volume changes in early RRMS: a 2-year longitudinal study. Neurology 2005;64:1001–7. [PubMed: 15781816]
24. Chard DT, Griffin CM, Rashid W, et al. Progressive grey matter atrophy in clinically early relapsing-remitting multiple sclerosis. Mult Scler 2004;10:387–91. [PubMed: 15327034]
25. Sormani MP, Rovaris M, Valsasina P, et al. Measurement error of two different techniques for brain atrophy assessment in multiple sclerosis. Neurology 2004;62:1432–4. [PubMed: 15111692]
26. FreeSurfer: a set of automated tools for reconstruction of the brain's cortical surface from structural MRI data, and overlay of functional MRI data onto the reconstructed surface. [12 September 2008]. Available at: http://surfer.nmr.mgh.harvard.edu

27. Hardmeier M, Wagenpfeil S, Freitag P, et al. Rate of brain atrophy in relapsing MS decreases during treatment with IFNβ-1a. Neurology 2005;64:236–40. [PubMed: 15668419]

28. Rudick RA, Fisher E, Lee JC, et al. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing–remitting MS. Neurology 1999;53:1698–704. [PubMed: 10563615]

Table 1

Baseline characteristics of the patients

	$RRMS_1$	$RRMS_2$	$RRMS_3$	Univariate p value
No of patients	116	26	109	
Sex (F/M)	83/33	18/8	85/24	0.81
Age (mean (SD))	34.1 (7.5)	36.7 (7.2)	34.9 (7.2)	0.14
Disease duration (years from first symptom) (mean (SD))	4.9 (3.8)	10.4 (7.8)	4.7 (4.0)	0.002
EDSS score (median (interquartile range))	2 (1.5, 3.5)	1.5 (1, 2.5)	2 (1.5, 3)	0.006
WMV (mm^3) (mean (SD))	868 423 (85 185)	827 068 (68 313)	856 305 (78 411)	0.021
GMV (mm^3) (mean (SD))	683 707 (70 808)	693 395 (65 936)	688 304 (70 588)	0.51
BPV (mm^3) (mean (SD))	1 552 130 (103 000)	1 520 463 (95 968)	1 544 609 (87 652)	0.14

Univariate p values are from Wilcoxon tests comparing $RRMS_1$ and $RRMS_2$; similar p values were obtained from the three group comparison.

BPV, brain parenchymal volume; EDSS, expanded disability status scale; GMV, gray matter volume; RRMS, relapsing–remitting multiple sclerosis; $RRMS_1$, untreated cohort of RRMS patients observed monthly for 9 months[14]; $RRMS_2$, untreated cohort of RRMS patients observed every 3 months for 1.5 years[17]; $RRMS_3$, treated cohort of RRMS patients observed monthly for 9 months[14]; WMV, white matter volume.

Table 2

Volume change estimates from the mixed effects model

	$RRMS_1$	$RRMS_2$	$RRMS_3$
GMV			
Annualised % change	−3.57	−1.89	−3.15
σ_{ri}	0.11	0.078	0.10
σ_{rs}	0.037	0.040	0.036
ρ_r	−0.13	−0.49	0.24
σ_e	0.023	0.029	0.023
WMV			
Annualised % change	1.05	−0.40	1.46
σ_{ri}	0.093	0.065	0.086
σ_{rs}	0.035	0.028	0.028
ρ_r	0.12	−0.07	0.11
σ_e	0.022	0.028	0.022
BPV			
Annualised % change	−0.94	−1.04	−0.53
σ_{ri}	0.071	0.063	0.057
σ_{rs}	0.013	0.015	0.014
ρ_r	−0.17	−0.38	0.12
σ_e	0.0080	0.014	0.0078

Estimated per cent change and variance components for GMV, WMV and BPV from each dataset.

σ_{ri} is the SD of the random intercept; σ_{rs} is the SD of the random slope; ρ_r is the correlation between the random intercept and slope; and σ_e is the residual SD. BPV, brain parenchymal volume; GMV, gray matter volume; RRMS, relapsing–remitting multiple sclerosis; $RRMS_1$, untreated cohort of RRMS patients observed monthly for 9 months[14]; $RRMS_2$, untreated cohort of RRMS patients observed every 3 months for 1.5 years[17]; $RRMS_3$, treated cohort of RRMS patients observed monthly for 9 months[14]; WMV, white matter volume.

Table 3

Sample size required to detect the specified reduction in the per cent change estimated from each dataset (table 2) based on a 2 year study with a baseline, year 1 and year 2 scan

Dataset	Measurement	33% reduction	50% reduction	90% reduction
$RRMS_1$	GMV	328	147	46
	WMV	3859	1712	527
	BPV	603	269	84
$RRMS_2$	GMV	1127	503	157
	WMV	18944	8425	2605
	BPV	724	323	100
$RRMS_3$	GMV	419	187	59
	WMV	1414	627	193
	BPV	2391	1064	329

Each sample size calculation assumes a 1:1 allocation of patients to treatment and placebo, 80% power and a two sided significance level of 0.05. Sample sizes are totals for both arms.

BPV, brain parenchymal volume; GMV, gray matter volume; RRMS, relapsing–remitting multiple sclerosis; RRMS1, untreated cohort of RRMS patients observed monthly for 9 months[14]; RRMS2, untreated cohort of RRMS patients observed every 3 months for 1.5 years[17]; $RRMS_3$, treated cohort of RRMS patients observed monthly for 9 months[14]; WMV, white matter volume.

Table 4

Relative sample size required for trials with different numbers/spacing of scans or length of trials compared with a 2 year trial with a baseline, year 1 and year 2 scan presented in table 3

Trial design	GMV	WMV	BPV
Varying the number/spacing of scans			
Two scans: Baseline and year 2	1	1	1
Three scans: Baseline, month 22, year 2	0.969	0.969	0.968
Four scans: Baseline and every 8 months to year 2	0.986	0.983	0.984
Four scans: Baseline, month 2 month, month 22, year 2	0.942	0.928	0.935
Five scans: Baseline and every 6 months to year 2	0.972	0.965	0.969
Five scans: Baseline, month 2, month 20, month 22, year 2	0.933	0.919	0.926
Varying the length of trial			
Three scans: Baseline, month 6 and year 1	1.46	1.50	1.49
Three scans: Baseline, year 1.5 and year 3	0.917	0.905	0.910
Three scans: Baseline, year 2 and year 4	0.889	0.871	0.879

Only the results for the $RRMS_1$ dataset with per cent change from table 2 are presented but similar results were seen for the other datasets and the per cent change from the literature.[12]

BPV, brain parenchymal volume; GMV, gray matter volume; $RRMS_1$, untreated cohort of RRMS patients observed monthly for 9 months[14]; WMV, white matter volume.