Sample-size calculations for short-term proof-of-concept studies of tissue protection and repair in multiple sclerosis lesions via conventional clinical imaging

Daniel S. Reich, MD, PhD^{1,*}, Richard White, MSc², Irene C. M. Cortese, MD¹, Luisa Vuolo, MD³, Colin D. Shea, MS¹, Tassie L. Collins, PhD⁴, and John Petkau, PhD²

¹Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA

²Department of Statistics, University of British Columbia, Vancouver, Canada

³Departments of Neurology and Radiology, University of Florence, Italy

⁴Myelin Repair Foundation, Saratoga, California, USA

Abstract

Background—New MS lesion activity on MRI can test immunomodulatory therapies in proof-of-concept trials. Comparably powerful endpoints to assess tissue protection or repair are lacking.

Objective—To report sample-size calculations for assessment of new lesion recovery.

Methods—In 2 sets of 6 active MS cases, new lesions were observed by monthly MRI for approximately 12 months. Averages and quartiles of normalized (proton density/T1/T2 weighted) and quantitative (T1/T2 and mean diffusivity maps for dataset 1, T2 and magnetization transfer ratio maps for dataset 2) measures were used to compare the lesion area before lesion appearance to afterward. A linear mixed effects model incorporating lesion- and participant-specific random effects estimated average levels and variance components for sample-size calculations.

Results—In both datasets, greatest statistical sensitivity was observed for the 25th percentile of normalized proton density-weighted signal. At 3T, using new lesions $\geq 15 \text{ mm}^3$, as few as 9 participants/arm may be required for a 6-month placebo-controlled add-on trial postulating a therapeutic effect size of 20% and statistical power of 90%.

Conclusion—Lesion recovery is a powerful outcome measure for proof-of-concept clinical trials of tissue protection and repair in MS. The trial design requires active cases and is therefore best implemented near disease onset.

Introduction

The key pathological feature of multiple sclerosis (MS) is the focal white matter lesion, harboring inflammation, immune-mediated demyelination, gliosis, and axonal injury and loss (1). The appearance of such a lesion in an eloquent portion of the central nervous

* Address for correspondence: 10 Center Drive MSC 1400, Building 10 Room 5C103, Bethesda, MD 20892, USA.

system is the basis for clinical relapse, and remission from relapse most likely involves resolution of inflammation, plasticity, and some degree of remyelination (2). Axonal damage, on the other hand, occurring both within and outside acute and chronic lesions, underlies the permanent neurological deficits that accrue in the progressive phase of the disease (3).

Over the last two decades, nearly a dozen medications have been approved for relapsing-remitting MS, all of them immunomodulatory. Despite reducing relapse rate and formation of new lesions, none definitively promotes healing of damaged tissue or significantly modulates progressive disease. There is therefore great hope that medications conferring tissue protection, which for the purposes of this paper also subsumes repair and remyelination, will prevent axonal loss in the earliest phases of the disease, and, in so doing, will be effective for limiting the long-term impact of MS.

A major problem in the field is that there is no widely accepted proof-of-concept clinical trial design to effectively and rapidly test putative tissue-protective agents (4). This is largely due to the absence of robust, sensitive markers of tissue degeneration that change sufficiently rapidly – typically 6 months or less. In addition, it remains unclear whether such agents should be first tested in progressive MS, where relapses and their accompanying focal white matter lesions are relatively few, or in early MS, where they are common. If new lesions could be used to assess more global processes of tissue protection and repair, then early MS would clearly be preferred. However, with increasingly effective immunomodulatory therapies, new lesions are ever rarer, necessitating enrollment into such trials essentially at the time of diagnosis. Furthermore, tissue-protection trials at this stage of the disease would need to be performed on a background of standard-of-care immunomodulatory therapies, as tissue-protective therapies should not, on their own, reduce the incidence of new inflammation. Finally, it would still be necessary to determine whether compounds that show tissue-protective efficacy in the setting of acute inflammation will also protect at-risk brain tissue that degenerates slowly over time.

This report retrospectively marshals data from a natural history study of two sets of 6 MS cases to generate sample size calculations for future proof-of-concept clinical trials for lesion repair. Candidate lesion-based outcome measures were derived from the monthly MRI scans and subsequently simulated in a series of related short-term (4 to 12 month) clinical trials. The results demonstrate that the strategy is promising and should be further tested in additional natural history studies and clinical trials.

Methods

Full methods are presented in the supplement. Briefly, we invited 6 individuals with relapsing-remitting MS (5) to undergo monthly MRI based on high lesion activity on screening scans. Study procedures were approved by the Institutional Review Board. Three tesla (T) MRI included the following scans: 3D T1-weighted before and after intravenous administration of standard gadolinium contrast; 2D double-echo for proton-density (PD) and T2 weighting as well as T2 mapping; 3D T2-FLAIR; 3D T1 mapping; and diffusion tensor imaging (DTI).

A second set of 6 individuals with relapsing-remitting MS was identified from the research database at our center, based on availability of the following pulse sequences acquired monthly for 1 year on a 1.5T scanner: 3D T1-weighted, 2D double-echo for PD and T2 weighting, 2D T2-FLAIR, and post-contrast T1. Cases required at least 2 gadolinium-enhancing lesions over the course of the year. Magnetization transfer-weighted scans obtained at the same time were analyzed when available.

Images were rigidly aligned across time, and white matter lesions and normal brain structures were segmented. Following intensity normalization by creation of standard (z) scores relative to “normal-appearing” white matter (NAWM), new lesions were identified on subtraction images and tracked over time (see the supplement for more details on the normalization). Within each lesion at each time point, the average and 25th, 50th, and 75th percentiles of normalized (n) T1, T2, PD, and FLAIR signal intensities were recorded, as well as quantitative (q) estimates of T1, T2, mean diffusivity (MD), and magnetization transfer ratio (MTR).

Time evolution of the normalized and quantitative MRI measures was analyzed via a mixed-effects model schematized in Figure 1. In addition to the mean MRI signal levels depicted in Figure 1, the model incorporates lesion- and participant-specific random effects for both the level of the MRI signal and the change in the MRI signal from the pre-lesion appearance to the post-lesion appearance time period. This model estimates a lesion’s “lack of recovery” – the difference between stabilized MRI measures taken following some delay after the lesion’s first appearance and the same measures taken prior to the lesion’s appearance. The model is used to generate estimated sample sizes for future clinical trials.

Results

Table 1 shows that participants were 50 years old and carried low disability at baseline; no participant sustained increased disability over the observation period. Treatment with disease-modifying therapy was initiated in participants who were initially untreated, and in some cases treatment changed during the study due to continued symptoms and/or MRI activity. Attribution of changes in MRI measures to particular treatments is beyond the scope of this study.

Figures 2 and 3 show examples of lesions tracked over the course of the study in two different cases from the 3T dataset. Over time, the volume of the lesion depicted in Figure 2 diminished slightly, but the lesion itself remained visible on all MRI contrasts. The two lesions tracked in Figure 3 appeared at different time points, and the lesion volumes were roughly stable over time. However, compared to the lesion in Figure 2, there was greater variation in lesion conspicuity across MRI contrasts in this case, with consistently good visualization on nT1, nPD, and qT1; slightly poorer visualization over time on nT2; and generally poor visualization on qT2 and MD.

Figure 4 shows the time courses of lesion volume and median MRI measures for each lesion from participant 3T-4. For nPD, which provided optimal sample sizes (see below), time courses of the 25th and 75th percentiles of normalized signal intensity are also shown. For

nPD, nT1, and qT1, sharp changes occurred at the time of lesion appearance, but otherwise the values remained roughly stable over time (with considerable variation). For nT2 and qT2, the value at the time of lesion appearance is greatest, and there is partial resolution over approximately 2 months. For MD, there is little discernible pattern.

The data for the 5 3T cases with new lesions are summarized in Figure 5. For each lesion, averages and standard deviations are calculated across all time points prior to lesion appearance and across time intervals successively farther removed from the time of lesion appearance; these values are subsequently pooled across all lesions for that participant. Data obtained at the time of lesion appearance are excluded, as these may be influenced by transient intralesional and perilesional edema. Although MRI values gradually trend toward their prelesional levels, this trend is generally fairly limited, and no measure fully normalizes. There is clear heterogeneity across participants, particularly in the prelesional value for each MRI measure and in the typical extent of MRI change after lesion appearance. These features are also heterogeneous across lesions within participants (not shown). Interestingly, the standard deviations are also mostly stable over time intervals after lesion appearance and are typically somewhat larger than before lesion appearance; including all data subsequent to lesion appearance is associated with the greatest variability. Importantly, standard deviations are quite comparable across participants. Qualitatively similar patterns were observed with the 1.5T dataset.

The linear mixed-effects model incorporates the sources of variation described above and can be used to assess the sensitivity of all combinations of MRI measure, within-lesion measurement descriptor, and post-lesion-appearance time-interval exclusion. Sample-size factors (SSfactor; see supplemental methods) derived from this model are presented in Table 2 for a trial involving participants similar to the study population of each set, assuming 12 months of treatment with monthly MRI scans. Under the usual assumption that the only impact of the test treatment is on the parameter of interest, in this case the mean lesion recovery (see the supplement for more detailed discussion of the assumptions), the required number of such participants per group is directly proportional to these sample size factors. SSfactor values were also evaluated when the analysis was restricted to lesions that enhanced on first appearance or to lesions ≥ 15 or $\geq 34 \text{ mm}^3$. Similar overall patterns were observed, and in all instances, nPD₂₅ including all scans subsequent to lesion appearance yielded the smallest SSfactor values.

The results in Table 2 can be summarized as follows:

- In each row, nPD has the smallest SSfactor. For the 3T and 1.5T sets, respectively, nT2 and nT1 generally have the next smallest SSfactor. (Note that since nT1 decreases in lesions, the 75th percentile for nT1 should be compared to the 25th for all of the other measures except for MTR, which also decreases in lesions.) In general, for the most relevant rows, where SSfactor values are small, nPD yields an approximately 30% smaller SSfactor than the next best measure.
- From a sample-size point of view, none of the quantitative measures is competitive.
- nPD_{ave} and nPD₅₀ have comparable SSfactor values, but those for nPD₂₅ are consistently smaller.

- Sensitivity generally decreases when additional time subsequent to lesion appearance is excluded. For statistical efficiency, therefore, the measures on all scans subsequent to lesion appearance should be included in the statistical analysis.
- Relative to the 3T dataset, the SSfactor for the 1.5T dataset is approximately 25% larger for the most efficient nPD measures, half the size for nT1, twice as large for nT2, and 3 times larger for qT2. Given the many differences between the 3T and 1.5T datasets, it is difficult to attribute these varying results to a single source.

SSfactor values for a trial with 6 months of treatment and monthly 3T MRI scans display a similar overall pattern (not shown). For this scenario, where the SSfactor values for nPD and nT2 are both small, the SSfactor for nT2 ranges from about 30% to more than 100% larger than that for nPD, so nT2 would again require a substantially larger sample size. As in Table 2, nPD₂₅ with all scans subsequent to lesion appearance included in the analysis yields the smallest SSfactor values.

As a real-world clinical trial would probably not include participants exactly similar to those reported in the 3T and 1.5T datasets, simulations were performed to obtain required sample sizes allowing for randomly differing numbers and patterns of new lesion appearance (see supplemental methods). Selected results for nPD₂₅, when all scans subsequent to lesion appearance are included, are shown in Figure 6 for the 3T dataset. The results indicate that even small effect sizes (~20%) are detectable in short studies with small numbers of participants (~9 per arm for a 6-month trial that considers recovery in lesions $\geq 15 \text{ mm}^3$ in volume, or roughly 3 mm in diameter). Reducing treatment duration from 12 months, even to as short as 4 months, increases the required sample size by at most 40–60%; this reflects the fact that, in this study population (as is typical in active MS cases), most new lesions appeared either at baseline or quite early during follow-up. Limiting the study to lesions that enhance with gadolinium at the time of first appearance has little effect. Excluding the smallest lesions ($< 15 \text{ mm}^3$) reduces the required sample size by more than 30% at each treatment duration considered, most likely due to increased observation-to-observation variation within smaller lesions. On the other hand, further restricting to lesions $\geq 34 \text{ mm}^3$ leads to some loss of sensitivity. Similar trends were observed from simulations based on the 1.5T dataset, although the gain due to excluding the smallest lesions was more modest and the increase in required sample size due to reducing the treatment duration was more substantial. This reflects the smaller proportion of new lesions that appeared during very early follow-up in the 1.5T dataset.

For a trial with 90% power, a two-sided 5% hypothesis test, a 20% postulated effect size, and lesions $\geq 15 \text{ mm}^3$, the required 3T sample sizes per arm (appropriately taking into account that some patients will contribute no or very few new lesions) are as follows: 10 (4 months), 9 (6 months), 8 (9 months), and 7 (12 months). The corresponding sample sizes at 1.5T are: 13 (4 months), 8 (6 months), 7 (9 months), and 7 (12 months). As a 4-month study could in principle lead to a relatively sparse dataset with few repeated observations over time on any single new lesion, making it difficult to fit our statistical model, a trial duration of at least 6 months seems reasonable.

Discussion

Our major conclusion is that a proof-of-concept clinical trial for repair of inflammatory, demyelinating, white matter lesions in MS can be performed quickly and efficiently, particularly at 3T (e.g., 9 participants per arm for 6 months of treatment). This represents a substantial improvement over currently employed trial designs for tissue protection/repair in MS using outcome measures of brain volume (6, 7) or retinal nerve fiber layer thickness (8). Of course, whether efficacy in promoting lesion recovery might predict efficacy for slowing disability accumulation in progressive MS, or even brain atrophy, remains to be determined.

New white matter lesions result from a complex cascade of events mediated by a variety of cell types and soluble factors (1). Serial imaging studies indicate that lesion growth and development is highly dynamic, tending to stabilize by about 6 months or even sooner (9–12). This dynamic behavior offers a unique opportunity to study processes of tissue damage and repair as they evolve within white matter lesions.

The incidence of new MS lesions is a powerful way to screen new immunomodulatory drugs for future Phase 3 clinical trials (13). Such exploratory studies often use a baseline-to-treatment design and a typical sample size of 10 to 15 participants. What is uncertain, and not typically tested in these trials, is whether lesions that form while participants are taking these drugs are different from those that form off treatment. To answer this question, a different kind of clinical trial is required.

As an outcome measure, recovery of new lesions by MRI is not a new idea. Evolution of new lesions to hypointense foci on T1-weighted spin-echo images (“black holes”) and within-lesion changes in MTR have been investigated (14–16), as has average MTR change across all white matter lesions (17). These measures were investigated in part because they were thought to be somewhat pathologically specific (black holes for irreversible tissue damage and MTR for demyelination), but despite years of study they have not garnered widespread use as primary outcome measures. One reason is sensitivity to scanning hardware and software, a problem that is compounded by large sample sizes that necessitate a multisite trial. At the same time, the pathological specificity of these MRI measures, particularly MTR (despite its known sensitivity to myelin), has repeatedly come into question (18–20). The approach described here deliberately changes the focus from specificity to sensitivity – that is, to an outcome measure and a trial design that can detect improvement in tissue damage rapidly and accurately, but that cannot on its own provide information about the biological mechanisms underlying that improvement.

Is it sensible to study tissue protection and repair in early MS? There is abundant evidence that at least some of the processes at play later in the disease, such as axonal loss, are relevant early on (21). However, relapses are generally followed by complete or near-complete functional recovery, probably due to endogenous remyelination and adaptive reorganization at the molecular, synaptic, and cellular levels. Nonetheless, most newly forming lesions persist on MRI for the duration of the disease, and white matter lesions accumulated during the 5 years following diagnosis correlate with disability 20 years later (22). It stands to reason, therefore, that a successful intervention that protects axons from the

insult of acute inflammation and demyelination would justify more expensive and arduous validation in clinical trials for progressive MS.

Our results identify nPD_{25} – the 25th percentile of the normalized proton density-weighted MRI signal – as the most sensitive outcome measure in short-term trials of tissue protection and repair in white matter lesions, which most likely would be conducted in combination with approved disease-modifying therapy. Five aspects of this outcome measure require elaboration: (1) use of normalized MRI signal rather than quantitative measures; (2) implications of PD weighting; (3) extraction of the 25th percentile of within-lesion signal intensity; (4) requirement for prelesional measurements; and (5) differences between the 1.5T and 3T sets.

Quantification of the MRI signal is attractive because the resulting maps are interpretable in physical terms and account for some of the technical issues that arise in image acquisition. Quantification should therefore allow comparisons across time and imaging site, facilitating multicenter trials. However, generation of quantitative maps – here, $qT1$, $qT2$, MD, and MTR – requires integration of multiple image types, each with associated noise, and the resulting error propagation mitigates the benefits of quantification. On the other hand, direct signal intensity normalization, required for comparison of conventional images across time, space, and person, is itself not straightforward. Our normalization procedure – creation of standard scores relative to NAWM – is based on tissue segmentation that does not need to be especially precise, as the required statistical moments of NAWM signal intensity are robustly estimated. This same normalization procedure has proved useful in other applications (23, 24) and importantly requires only a single image volume, eliminating the problem of error propagation.

PD-weighted images are commonly used in MS to segment white matter lesions (25) and are routinely recommended for diagnostic assessment (26), but they are not commonly applied to assessing lesional changes over time. However, tissue damage, including demyelination, can affect PD (19, 27), so it is plausible that nPD will be useful for tracking lesional changes over time in a clinical trial setting. To limit the influence of resolving edema on nPD measures, data from the scan obtained at the time of each lesion's first appearance were excluded from the statistical analysis; exclusion of additional scans beyond the first appears to be unnecessary.

For each MRI measure, the evolution of the median and average as well as the 25th and 75th percentiles of the within-lesion measurements was studied. Median and average performed similarly. Examination of quartiles was based on the idea that whereas tissue damage is most severe at the center of a newly developing lesion (28, 29), tissue repair, including remyelination, may be partial (30, 31). Thus, quartiles might be especially sensitive to differences between severely damaged and partially repaired tissue while avoiding precise modeling of spatial changes, particularly near edges of developing and repairing lesions, where partial volume averaging with surrounding extralesional tissue may be problematic (32).

Prelesional MRI measures are required to calculate a lesion's "lack of recovery," as defined here. As lesion volume may change over time, it is difficult to precisely define the prelesional area. We investigated a number of possible solutions, ultimately selecting the lesion mask from the first time point at which the lesion was visible and propagating it backward in time. (An alternative analysis, in which measurements were made across all voxels that were ever included within the lesion, proved less sensitive.)

Finally, the overall conclusion that nPD_{25} is the optimal outcome measure was consistent across the two datasets, despite the fact that they were obtained on different MRI scanners at two different magnetic field strengths. Slight differences between the two datasets may reflect participant-specific differences as well as technical factors, including the signal-to-noise ratio of the scanner and head coil and the inplane resolution of the PD scan.

The sensitivity of this trial design arises from its control of multiple sources of variation; assessment of the effect of treatment is based on the within-lesion (not only within-participant) changes in the MRI signal from the pre-lesion time period to the post-lesion appearance time period. As always, required sample sizes are dependent on assumptions about the nature of the data that will be observed in the planned trial, as discussed in more detail in the supplement. Chief among these are the usual assumptions: (1) that the variance components determined from our study participants will be relevant for the participants on both arms of the planned trial; and (2) that the only impact of the test treatment is on the parameter of interest, in this case the mean lesion recovery.

For the trial design described here to be successful, a number of conditions are necessary – chiefly, observation of an adequate number of lesions. To keep sample sizes small, this condition necessitates enriching for study participants who develop new lesions. The optimal strategy for such enrichment has yet to be defined, and if trial participants have substantially fewer lesions than those described here, required sample sizes would increase. On the other hand, if lesions were more evenly distributed across the study population, fewer lesions in total would need to be observed.

This report is limited by the small number of participants, although the similarity of the results in both datasets is encouraging. Replication of the finding in additional datasets would be valuable. Moreover, although we investigated MTR in the 1.5T dataset, MTR data were missing in 45% of lesion observations. MTR is the quantitative MRI measure that is most commonly used to study remyelination in MS, and there is an extensive literature on lesion MTR as an outcome measure (16, 33, 34). However, MTR is difficult to standardize, and its calculation is subject to error propagation, as discussed above. Further work directly comparing MTR to normalized signal intensity measures such as nPD_{25} – perhaps in the context of post-hoc analysis of existing clinical trial datasets – would be informative.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Neuroimmunology Clinic (in particular, Joan Ohayon, Kaylan Fenton, and Helen Griffith) for expert clinical evaluation and care of the study participants. Govind Nair, Roger Stone, Ilena George, Iordanis Evangelou, Joelle Sarlls, Justin Senseney, John Ostuni, and the staff and technologists of the NIH Functional Magnetic Resonance Facility assisted with data acquisition and analysis. Issues related to image processing and normalization benefitted from discussions with Ciprian Crainiceanu, Taki Shinohara, and Elizabeth Sweeney. Henry McFarland, Peter Calabresi, and Scott Zamvil provided especially useful advice at several stages of the work. Research was supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke and the Myelin Repair Foundation.

References

1. Lassmann H. Review: the architecture of inflammatory demyelinating lesions: implications for studies on pathogenesis. *Neuropathol Appl Neurobiol.* 2011; 37(7):698–710. [PubMed: 21696413]
2. Tomassini V, Matthews PM, Thompson AJ, Fuglo D, Geurts JJ, Johansen-Berg H, et al. Neuroplasticity and functional recovery in multiple sclerosis. *Nature reviews Neurology.* 2012
3. Trapp BD, Ransohoff R, Rudick R. Axonal pathology in multiple sclerosis: relationship to neurologic disability. *Curr Opin Neurol.* 1999; 12(3):295–302. [PubMed: 10499174]
4. Barkhof F, Calabresi PA, Miller DH, Reingold SC. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. *Nature reviews Neurology.* 2009; 5(5):256–66. [PubMed: 19488083]
5. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria”. *Ann Neurol.* 2005; 58(6):840–6. [PubMed: 16283615]
6. Altmann DR, Jasperse B, Barkhof F, Beckmann K, Filippi M, Kappos LD, et al. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. *Neurology.* 2009; 72(7):595–601. [PubMed: 19005170]
7. Healy B, Valsasina P, Filippi M, Bakshi R. Sample size requirements for treatment effects using gray matter, white matter and whole brain volume in relapsing-remitting multiple sclerosis. *Journal of neurology, neurosurgery, and psychiatry.* 2009; 80(11):1218–23.
8. Henderson AP, Altmann DR, Trip AS, Kallis C, Jones SJ, Schlottmann PG, et al. A serial study of retinal changes following optic neuritis with sample size estimates for acute neuroprotection trials. *Brain.* 2010; 133(9):2592–602. [PubMed: 20566483]
9. Meier DS, Guttmann CR. Time-series analysis of MRI intensity patterns in multiple sclerosis. *NeuroImage.* 2003; 20(2):1193–209. [PubMed: 14568488]
10. Guttmann CR, Ahn SS, Hsu L, Kikinis R, Jolesz FA. The evolution of multiple sclerosis lesions on serial MR. *AJNR Am J Neuroradiol.* 1995; 16(7):1481–91. [PubMed: 7484637]
11. Gaitan MI, Shea CD, Evangelou IE, Stone RD, Fenton KM, Bielekova B, et al. Evolution of the blood-brain barrier in newly forming multiple sclerosis lesions. *Ann Neurol.* 2011; 70(1):22–9. [PubMed: 21710622]
12. van Waesberghe JH, van Walderveen MA, Castelijns JA, Scheltens P, Lycklama a Nijeholt GJ, Polman CH, et al. Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin-echo and magnetization transfer MR. *AJNR Am J Neuroradiol.* 1998; 19(4):675–83. [PubMed: 9576653]
13. McFarland HF, Stone LA, Calabresi PA, Maloni H, Bash CN, Frank JA. MRI studies of multiple sclerosis: implications for the natural history of the disease and for monitoring effectiveness of experimental therapies. *Mult Scler.* 1996; 2(4):198–205. [PubMed: 9345374]
14. van den Elskamp IJ, Lemcke J, Dattola V, Beckmann K, Pohl C, Hong W, et al. Persistent T1 hypointensity as an MRI marker for treatment efficacy in multiple sclerosis. *Mult Scler.* 2008; 14(6):764–9. [PubMed: 18611989]
15. Richert ND, Ostuni JL, Bash CN, Leist TP, McFarland HF, Frank JA. Interferon beta-1b and intravenous methylprednisolone promote lesion recovery in multiple sclerosis. *Mult Scler.* 2001; 7(1):49–58. [PubMed: 11321194]

16. Brown RA, Narayanan S, Arnold DL. Segmentation of magnetization transfer ratio lesions for longitudinal analysis of demyelination and remyelination in multiple sclerosis. *Neuroimage*. 2012; 66C:103–9. [PubMed: 23110887]
17. Altmann DR, Button T, Schmierer K, Hunter K, Tozer DJ, Wheeler-Kingshott CA, et al. Sample sizes for lesion magnetisation transfer ratio outcomes in remyelination trials for multiple sclerosis. *Multiple Sclerosis and Related Disorders*. 2013 in press.
18. Vavasour IM, Laule C, Li DK, Traboulsee AL, MacKay AL. Is the magnetization transfer ratio a marker for myelin in multiple sclerosis? *J Magn Reson Imaging*. 2011; 33(3):713–8. [PubMed: 21563257]
19. Mottershead JP, Schmierer K, Clemence M, Thornton JS, Scaravilli F, Barker GJ, et al. High field MRI correlates of myelin content and axonal density in multiple sclerosis—a post-mortem study of the spinal cord. *J Neurol*. 2003; 250(11):1293–301. [PubMed: 14648144]
20. Schmierer K, Altmann DR, Kassim N, Kitzler H, Kerskens CM, Doege CA, et al. Progressive change in primary progressive multiple sclerosis normal-appearing white matter: a serial diffusion magnetic resonance imaging study. *Mult Scler*. 2004; 10(2):182–7. [PubMed: 15124765]
21. Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mork S, Bo L. Axonal transection in the lesions of multiple sclerosis. *N Engl J Med*. 1998; 338(5):278–85. [PubMed: 9445407]
22. Fisniku LK, Brex PA, Altmann DR, Miszkiet KA, Benton CE, Lanyon R, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain*. 2008; 131(Pt 3):808–17. [PubMed: 18234696]
23. Sweeney EM, Shinohara RT, Shea CD, Reich DS, Crainiceanu CM. Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal MRI. *AJNR Am J Neuroradiol*. 2013; 34(1):68–73. [PubMed: 22766673]
24. Shinohara RT, Crainiceanu CM, Caffo BS, Gaitan MI, Reich DS. Population-wide principal component-based quantification of blood-brain-barrier dynamics in multiple sclerosis. *Neuroimage*. 2011; 57(4):1430–46. [PubMed: 21635955]
25. Horsfield MA, Bakshi R, Rovaris M, Rocca MA, Dandamudi VS, Valsasina P, et al. Incorporating domain knowledge into the fuzzy connectedness framework: application to brain lesion volume estimation in multiple sclerosis. *IEEE Trans Med Imaging*. 2007; 26(12):1670–80. [PubMed: 18092737]
26. Simon JH, Li D, Traboulsee A, Coyle PK, Arnold DL, Barkhof F, et al. Standardized MR imaging protocol for multiple sclerosis: Consortium of MS Centers consensus guidelines. *AJNR Am J Neuroradiol*. 2006; 27(2):455–61. [PubMed: 16484429]
27. Nijeholt GJ, Bergers E, Kamphorst W, Bot J, Nicolay K, Castelijns JA, et al. Post-mortem high-resolution MRI of the spinal cord in multiple sclerosis: a correlative study with conventional MRI, histopathology and clinical phenotype. *Brain*. 2001; 124(Pt 1):154–66. [PubMed: 11133795]
28. Henderson AP, Barnett MH, Parratt JD, Prineas JW. Multiple sclerosis: distribution of inflammatory cells in newly forming lesions. *Ann Neurol*. 2009; 66(6):739–53. [PubMed: 20035511]
29. Prineas JW, Parratt JD. Oligodendrocytes and the early multiple sclerosis lesion. *Ann Neurol*. 2012; 72(1):18–31. [PubMed: 22829266]
30. Schmierer K, Parkes HG, So PW. Direct visualization of remyelination in multiple sclerosis using T2-weighted high-field MRI. *Neurology*. 2009; 72(5):472. [PubMed: 19188581]
31. Patrikios P, Stadelmann C, Kutzelnigg A, Rauschka H, Schmidbauer M, Laursen H, et al. Remyelination is extensive in a subset of multiple sclerosis patients. *Brain*. 2006; 129(Pt 12):3165–72. [PubMed: 16921173]
32. Chen JT, Kuhlmann T, Jansen GH, Collins DL, Atkins HL, Freedman MS, et al. Voxel-based analysis of the evolution of magnetization transfer ratio to quantify remyelination and demyelination with histopathological validation in a multiple sclerosis lesion. *NeuroImage*. 2007; 36(4):1152–8. [PubMed: 17543541]
33. Arnold, DL.; Dalton, CM.; Schmierer, K.; Pike, GB.; Miller, DH. Imaging of demyelination and remyelination in multiple sclerosis. Duncan, ID.; Franklin, RJM., editors. New York: Springer; 2013. p. 233-53.

34. van den Elskamp IJ, Knol DL, Vrenken H, Karas G, Meijerman A, Filippi M, et al. Lesional magnetization transfer ratio: a feasible outcome for remyelinating treatment trials in multiple sclerosis. *Mult Scler*. 2010; 16(6):660–9. [PubMed: 20350960]

Figure 1.

Schematic depiction of the statistical model used for sample size determination, focusing on a single lesion that appears during the study (time point labeled “lesion appearance”). For simplicity, only a single pre-appearance scan (“baseline”) and a “final” scan are depicted. In the top row, gray boxes represent an area of normal-appearing white matter in which the lesion is located. At baseline, a dashed circle denotes the location where the lesion will appear. Beginning at the time of lesion appearance, the MRI signal intensity of the lesion is depicted by the brightness within the circle and by the curves in the bottom row. The model postulates that the final signal intensity is less abnormal in the treatment arm, and the difference in signal intensity between the two arms, averaged across all lesions and study participants, is the “treatment effect.” Note that data obtained around the time of lesion appearance are not analyzed due to the expected presence of edema at that time point.

Figure 2.

Time evolution of a lesion from participant 3T-3, a woman with relapsing-remitting MS who was 30 years old when first studied (“month-2”). A, images taken at month 2, when two contrast-enhancing lesions appeared (arrows in A1). A1, T1-weighted post contrast; A2, T2-FLAIR post-contrast; A3, proton density (PD)-weighted pre-contrast. The red boxes outline the regions enlarged in B, which shows the evolution of the larger lesion (white arrow in A1) over time. Image contrasts in B include normalized (n) T1, PD, and T2, quantitative (q) T1 and T2, and mean diffusivity (MD). The lesion mask, derived automatically from T1, PD, T2, and T2-FLAIR images, is shown in the top row for each time point. MD images were missing or of poor quality at months 3 and 9. Window and level settings were kept constant across time for each image type to highlight the stability of the normalization process.

Figure 3.

Time evolution of two lesions from participant 3T-4, a man with relapsing-remitting MS who was 28 years old when first studied (“month-5”). A, images taken at month 3, which show one contrast-enhancing (orange arrow in A1) and one non-enhancing lesion (white arrow); the non-enhancing lesion first appeared at month 0 but did not enhance on that scan. A1, T1-weighted post contrast; A2, T2-FLAIR post-contrast; A3, proton density (PD)-weighted pre-contrast. The red boxes outline the regions enlarged in B, which shows the evolution of both lesions over time. Image contrasts in B include T1-weighted post contrast, T2-FLAIR, (n) T1, PD, and T2, quantitative (q) T1 and T2, and mean diffusivity (MD). The lesion mask, derived automatically from T1, PD, T2, and T2-FLAIR images, is shown in the third row of B for each time point. The qT1 map was missing at time point -5 months. Window and level settings were kept constant across time for the normalized and quantitative images to highlight their stability. Note that the lesions become less conspicuous over time on some images (especially T2-FLAIR, nT2, qT2, and MD) but remain clearly visible on others (nT1, nPD, qT1).

Figure 4.

Time courses of all individual newly appearing lesions from participant 3T-4, a representative slice from whose MRI is shown in Figure 2. Each line represents a different lesion, centered around the time at which the lesion was first seen on MRI (“time 0”). For individual measurements, a prefix of “n” indicates “normalized,” in which the signal intensity is given in units of standard deviations of the same measurement across extralesional (“normal-appearing”) white matter; a prefix of “q” indicates “quantitative,” in which the measurement is an estimate of the T1 or T2 relaxation time. Subscripts indicate the percentile of the lesion-based measurement (25th, 50th, or 75th). PD, proton density. MD, mean diffusivity (from diffusion tensor imaging).

Figure 5.

Average (solid lines) and standard deviation (dashed lines) of MRI measurements across all lesions from each participant in the 3T dataset. Prior to summarizing across lesions, measurements are first considered within specific time windows: “Pre” indicates all time points prior to lesion appearance; “>0” indicates all time points after lesion appearance, excluding the first scan on which the lesion appeared; and “>45,” “>75,” and “>115” exclude measurements made prior to 45, 75, and 115 days, respectively, after lesion appearance. For individual measurements, a prefix of “n” indicates “normalized,” in which the signal intensity is given in units of standard deviations of the same measurement across extralesional (“normal-appearing”) white matter; a prefix of “q” indicates “quantitative,” in which the measurement is an estimate of the T1 or T2 relaxation time. Subscripts indicate the percentile of the lesion-based measurement (25th, 50th, or 75th). PD, proton density. MD, mean diffusivity (from diffusion tensor imaging).

Figure 6.

Required sample sizes (number of patients per arm for a trial with 1:1 randomization to test treatment and comparator) for nPD₂₅, the lesion-based MRI measurement that demonstrated the greatest statistical sensitivity as shown in Table 2. The left panel shows the required sample size as a function of the % treatment-related improvement in lesion “lack of recovery” for 12 months of treatment under four scenarios: all lesions, gadolinium-enhancing lesions only, lesions with volume $\geq 15 \text{ mm}^3$, and lesions with volume $\geq 34 \text{ mm}^3$. The right panel shows the same four scenarios as a function of trial duration for a fixed 20% treatment-related improvement in lesion lack of recovery. All values are based on the 3T dataset and are calculated for a two-sided test with $\alpha=0.05$ and 90% statistical power.

Table 1

Demographics

Participant	Sex	Age (years) at first scan	Years of symptoms at first scan	EDSS at first (last) scan	Treatment at baseline	New lesions at baseline *	Total lesions over 12 months *
3T-1	male	39	4	2.5 (2)	untreated	2/2/2/2	4/4/4/4
3T-2	male	40	8	2.5 (2)	untreated	2/2/2/2	6/6/6/6
3T-3	female	30	3	1 (2)	untreated	5/3/4/3	41/18/32/24
3T-4	male	28	0.3	1 (1)	subcutaneous interferon- β 1a	15/0/13/10	20/4/18/15
3T-5	male	25	2	1 (1)	glatiramer acetate	10/1/9/7	37/17/31/25
3T-6	male	48	8	1.5 (1)	glatiramer acetate	0/0/0/0	0/0/0/0
1.5T-1	male	18	4	1.5 (1.5)	subcutaneous interferon- β 1b	8/7/7/4	61/46/48/33
1.5T-2	female	29	???	1.5 (1)	untreated	0/0/0/0	37/32/33/28
1.5T-3	male	41	1	?? (1)	untreated	0/0/0/0	4/4/3/2
1.5T-4	male	36	???	1.5 (1)	untreated	1/1/1/0	3/2/2/2
1.5T-5	female	37	???	1.5 (1)	untreated	4/4/2/2	20/13/17/10
1.5T-6	male	29	0.3	1.5 (1)	untreated	8/8/8/5	17/17/16/18

* From left to right: total/gadolinium-enhancing/volume 15 mm³/volume 34 mm³.

Table 2

Sensitivity of MRI measures in a 12-month trial with monthly scanning.

Version of measure	Inclusion interval (days post lesion appearance)	Sample size factor*					
		nPD	nT1	nT2	qT1	qT2	MD
25 th percentile	>0	0.013	0.038	0.017	0.047	0.045	N/A
	>45	0.015	0.038	0.018	0.048	0.053	0.240
	>75	0.019	0.040	0.022	0.056	0.069	0.300
	>115	0.024	0.041	0.025	0.068	0.078	0.370
50 th percentile	>0	0.017	0.064	0.019	0.037	0.058	0.170
	>45	0.018	0.052	0.021	0.039	0.056	0.160
	>75	0.022	0.063	0.025	0.049	0.069	0.220
	>115	0.027	0.058	0.028	0.059	0.069	0.250
75 th percentile	>0	0.026	0.101	0.029	0.046	0.116	0.160
	>45	0.026	0.065	0.029	0.050	0.093	0.170
	>75	0.031	0.093	0.033	0.063	0.102	0.210
	>115	0.035	0.082	0.036	0.086	0.100	0.260
average	>0	0.016	0.052	0.021	0.032	0.079	0.131
	>45	0.018	0.039	0.022	0.038	0.094	0.154
	>75	0.021	0.047	0.026	0.048	0.114	0.205
	>115	0.026	0.045	0.029	0.066	0.144	0.251
25 th percentile	>0	0.017	0.045	0.032	0.032	0.224	0.179
	>45	0.019	0.049	0.037	0.030	0.207	0.150
	>75	0.019	0.048	0.037	0.030	0.143	0.198
	>115	0.018	0.056	0.034	0.031	0.215	0.306
50 th percentile	>0	0.020	0.029	0.041	0.042	0.392	0.179
	>45	0.020	0.031	0.045	0.041	0.366	0.178
	>75	0.020	0.031	0.045	0.040	0.169	0.234

Version of measure	Inclusion interval (days post lesion appearance)	Sample size factor*					
		nPD	nT1	nT2	qT1	qT2	MD
75 th percentile	>115	0.021	0.036	0.046	0.038	0.267	0.427
	>0	0.027	0.026	0.067	0.059	0.312	0.176
	>45	0.028	0.028	0.073	0.057	0.188	0.161
	>75	0.027	0.028	0.071	0.054	0.209	0.271
	>115	0.028	0.027	0.076	0.052	0.155	0.372
average	>0	0.020	0.031	0.048	0.042	0.233	0.181
	>45	0.021	0.033	0.053	0.040	0.252	0.154
	>75	0.021	0.032	0.053	0.038	0.314	0.222
	>115	0.021	0.034	0.054	0.036	0.150	0.363

* Exact expression provided in supplemental methods.