3.3

Lesional Magnetization Transfer Ratio — a Feasible Outcome for Remyelinating Treatment Trials in MS

I.J. van den Elskamp
D.L. Knol, H. Vrenken
G. Karas
A. Meijerman
M. Filippi
L. Kappos
F. Fazekas
K. Wagner
C. Pohl
R. Sandbrink
C.H. Polman
B.M.J. Uitdehaag
F. Barkhof

In press in Mult Scler 2010

ABSTRACT

Background: Magnetization transfer ratio (MTR) is a sensitive parameter to quantify the integrity of myelinated white matter in patients with MS. Lesional MTR decreases in the acute phase due to demyelination, and subsequently shows recovery depending on the degree of remyelination (in the absence of axonal loss). The recovery of average lesion MTR therefore, might prove a viable outcome measure to assess the effect of remyelinating agents.

Objective: To determine the required sample size for phase II multicentre clinical trials using the recovery of average lesion MTR as primary outcome measure.

Methods: With 7 monthly MRI scans, the MTR evolution of 349 new enhancing lesions before and after enhancement was assessed in 32 MS patients from 5 centres. Multilevel models were fitted to the data yielding estimates for the variance components which were applied in power calculations. Sample sizes were determined for placebo-controlled, multicentre trials using lesional MTR recovery post-enhancement as primary outcome measure.

Results: Average lesion MTR decreased slightly in the build-up to enhancement, decreased dramatically during enhancement and showed recovery in the period after cessation. The power calculations showed that for a power of 80%, approximately 160 patients per trial (mean number of 4 lesions per patient) are required to detect a 30% increase in lesional MTR post-enhancement compared to placebo, whereas 60 subjects are required to detect a 50% increase in lesional MTR compared to placebo, assuming a mild variation in the treatment-response between patients.

Conclusion*:* Recovery of lesion MTR is a feasible outcome measure for future multicentre clinical trials measuring the effect of remyelinating agents.

INTRODUCTION

White matter lesion formation is a hallmark in the pathology of multiple sclerosis (MS), and in the early phase of lesion development, lesions consist of focal inflammation accompanied by myelin destruction [1]. To quantify the extent of residual damage within lesions, magnetization transfer imaging is an effective MRI technique. This technique is based on the exchange of magnetization between protons in a restricted environment and those where motion is relatively free and, in the brain, largely driven by the integrity of myelin. In undisrupted white matter, the magnetic transfer ratio (MTR) is high due to the bound protons within myelin, whereas in MS lesions, demyelination causes an increase of the unbound protons and a significant decrease in MTR [2].

Longitudinal studies of MS lesions have revealed a characteristic pattern of MTR changes during lesion development. In the months preceding Gadolinium (Gd) enhancement, MTR slightly declines in the corresponding normal appearing white matter (NAWM; white matter not containing a lesion) where the lesion is to appear [3,4]. At the time of enhancement, the MTR decreases dramatically due to inflammation and demyelination [5] and in subsequent months, partial or complete MTR recovery may occur, consistent with repair which appears mainly driven by remyelination of axons which have remained intact [5,6]. Because of the greater pathological specificity for myelin content compared to conventional MRI [2], and the correlation with clinical disability [7], lesional MTR recovery is a promising outcome measure for clinical trials assessing the effect of therapies inducing remyelination.

At present, lesional MTR recovery post-enhancement has sporadically been applied as secondary outcome measure in clinical trials. One study compared the rate of increase of MTR after cessation of enhancement in new lesions before treatment with those after treatment with Interferon-beta1a (IFNb-1a), and showed a significantly faster rate of MTR recovery for lesions during therapy [8]. Another study showed the percentage MTR recovery of Gd enhancing lesions in patients before and after treatment with Interferon-beta1b to be significantly higher in Gd lesions during treatment [9]. Although these studies illustrate the potential of lesional MTR recovery as outcome measure in clinical trials, the results were obtained with small numbers of patients at single centres. If lesional MTR recovery is to be adopted as primary outcome measure in larger, multicentre phase II clinical trials of remyelinating agents

for MS, the required sample sizes to reliably demonstrate a treatment effect should be determined.

In this study, based on data acquired in a multicentre sample of RRMS patients without effective treatment, the MTR evolution of new enhancing lesions is assessed. Power calculations based on multilevel modelling are applied to determine the required sample size for placebo controlled clinical trials using lesional MTR recovery as primary outcome measure.

MATERIAL & METHODS

Patients

The MTR analyses were performed in a subcohort of 32 Relapsing Remitting MS (RRMS) patients in 5 centres from the placebo-controlled oral IFNb-1a study. Its study population is regarded as a natural history cohort since no clinical or MRI effect could be observed for any of the doses tested [10]. Patients in the current subcohort were randomly selected from the participating centres (original protocol and protocol amendments after study initiation were approved by ethics committees of participating centers before starting the study. Patients provided written informed consent).

MRI

For each patient, a series of seven MRI scans was performed at monthly intervals. In addition to the conventional imaging protocol [10], images for mapping MTR were acquired by performing two 2D gradient echo sequences (TR: 23 ms, TE: 4 ms, flip angle 20°, pixelsize: 1x1 mm, slice thickness: 5 mm) one with and one without a gaussian MT-prepulse (duration: 7.68 ms, offset 1500 Hz, equivalent flip angle 500°). Both sets were then co-registered, and the MTR was calculated for each pixel as the relative signal intensity decrease due to the application of the MT prepulse by:

$$\textbf{(eq. 1)} \quad MTR = \frac{M(0) - M(s)}{M(0)} \text{X}100\%$$

where M(0) indicates the signal intensity without the prepulse, and M(s) the intensity with the prepulse.

The MTR values of newly enhancing lesions were determined by the following steps. First, a radiologist marked all gadolinium-enhancing lesions on the T1 post-contrast scan (G.K). These markings were subsequently used by a technician to manually outline these lesions using in-house developed software (Show-images), creating regions of interest (ROIs). Secondly, all scans from month 1 to month 6 were co-registered to the baseline scan using a rigid body registration method. Thirdly, the transformation matrices were used to co-register the MTR scans and ROIs allowing accurate determination of MTR values in the designated ROIs before and after Gd-enhancement. In this way, a lesion appearing *e.g.* at the month 4 scan can be assessed for MTR evolution four months "pre-enhancement" and 2 months "post-enhancement". Enhancing lesions appearing on the first scan were not analyzed due to the lack of distinction between persisting and new enhancement. To reduce measurement error, periventricular lesions not distinguishable from CSF, extremely small lesions (<8mm^2), infratentorial lesions and ring enhancing lesions were not included in the analysis.

In addition to lesion MTR measurements, the MTR of NAWM was analyzed at all seven time points as a reference. In each patient, 2 to 4 ROIs were placed in the NAWM of the left and right frontal lobe on the last T2-weighted scan, and subsequently measured on the previous time-points, after verifying the absence of lesions in the ROIs at all time-points.

Statistical analysis

Multilevel models, assuming the repeated MTR measurements to be correlated within a lesion, the measurements within lesions to be correlated within patients, and the measurements within patients to be correlated within centres, were fitted to the NAWM and lesional MTR data to estimate the fixed effects and the components of variance of the individual levels (lesion, patient, centre and residual variance) (MLwiN v2.02). First, NAWM data was modelled, assuming a linear relation with time and was allowed random intercepts within patients and within centres. Second, the model for lesional MTR data describes the series of MTR measurements pre-enhancement (time -6 to time -1) and the series post enhancement (time 0 to time +5) by individual fixed slopes, and the difference between the first MTR measurement pre-enhancement and the last MTR measurement post-enhancement is allowed for using random slopes and random intercepts within lesions, within patients and within centres, yielding an overall model encompassing the heterogeneity of lesional MTR measurements. Thus, although a treatment effect is based on the first and last observation only,

both observations are greatly defined by the slope of the curve as determined with the intervening timepoints. The application of complexer models, including time as a factor, did not improve the fit.

Then, in analogy with Snijders **[11]** the formula for determining the required sample size for a given power, effect size and significance level (derived from the basic sample size formula) to accustom the four level structure of our data is determined (**Appendix)**. The definitive formula however, encompasses the occurrence of *interaction* between treatment and patients and between treatment and centres, (e.g. the response to treatment might vary between subjects or between centres) which cannot be estimated with the current data, since it requires MTR data with the presence of a genuine treatment which is unavailable at this moment. Therefore, we assume *no* interaction between treatment and centres in the present calculations, and simulate interactions of varying magnitude between treatment and patients to study their effect on sample size.

Sample sizes are estimated for multicentre clinical trials measuring lesional MTR of new enhancing lesions from 6 months pre enhancement to 5 months post enhancement. A treatment effect is expressed as the percentage reduction in residual MTR decrease, thus: the difference in percentage increase in MTR between the last MTR measurement post enhancement and the first MTR measurement pre enhancement in treated patients compared to untreated patients (**Figure 1**). Calculations are as aforementioned performed for trials with varying magnitude of interaction between treatment and patients, and for a varying mean numbers of lesions per patient, to assess the effect on the sample sizes of both sources of variation. All calculations are based on a 80% power to detect treatment effects, a 5% two-sided significance level and the absence of drop-outs and inactive patients (patients developing no enhancing lesions during follow-up). Depending on the expected percentage inactivity (percentage of patients developing no enhancing lesions at study end) and percentage dropout rate, current sample size can be manually adjusted by multiplying the given sample sizes by:

(**eq.2**) $$\frac{1}{(1 - p_{dropout})(1 - p_{inactive})}$$

where p_{droput} is the expected proportion drop-out, and $p_{inactive}$ the expected proportion of inactivity.

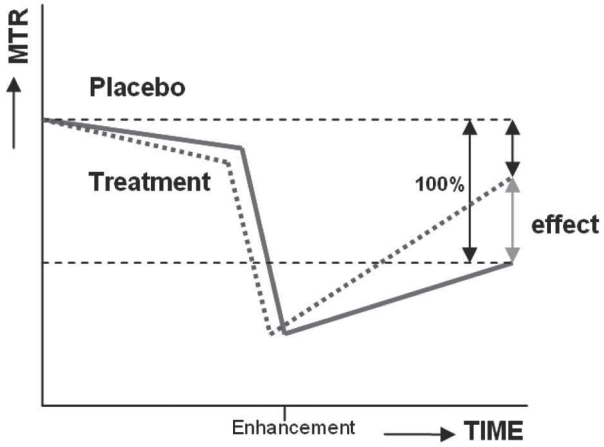

Figure 1 | Visualisation of the simulated treatment effect. The decrease in MTR in the placebo group between the last MTR measurement post enhancement and the first MTR measurement pre enhancement (line) is compared with the decrease in MTR in the treatment group (dotted), and the effect specified as the percentage increase in MTR in the treatment group compared to the placebogroup.

RESULTS

Demographics

Baseline demographics and MRI characteristics are shown in Table 1. No statistical differences were found between the current cohort and the remaining patients of the original study cohort (data not shown) [10].

MTR evolution

In total, 105 areas of NAWM were analyzed. The mean MTR course of the NAWM areas varies per centre, reflecting differences in MTR measurement technique per centre (Figure 2). The mean NAWM MTR varied between 36.3 (SD=1.5) and 30.6 (SD=4.0), with an overall NAWM MTR accross centres of 34.6 (SD=2.3). Multilevel modelling of the NAWM data showed that almost half of the total NAWM variance (7.84) consisted of the variance introduced by centres (3.86). As expected, the analysis showed the NAWM MTR measurements to be clustered within centres (X^2= 7.3, df=1, P=0.01).

Table 1 | Baseline demographics and MRI characteristics.

Characteristic	Value
General	
n	32
Sex (Female / Male)	20 / 12
age (Mean, SD)	32.3 (9.5)
disease duration in years (Mean, SD)	5.8 (6.3)
Baseline EDSS (Median, IQR)	2.0 (1.5-3.0)
MRI	
Number of Gd enhancing lesions (Median, SD)	1.5 (4.4)
Number of patients with ≥1 T1 Gd enhancing lesion (%)	24 (75%)

SD = Standard Deviation
IQR = Inter Quartile Range
EDSS = Expanded Disability Status Scale
Gd = Gadolinium

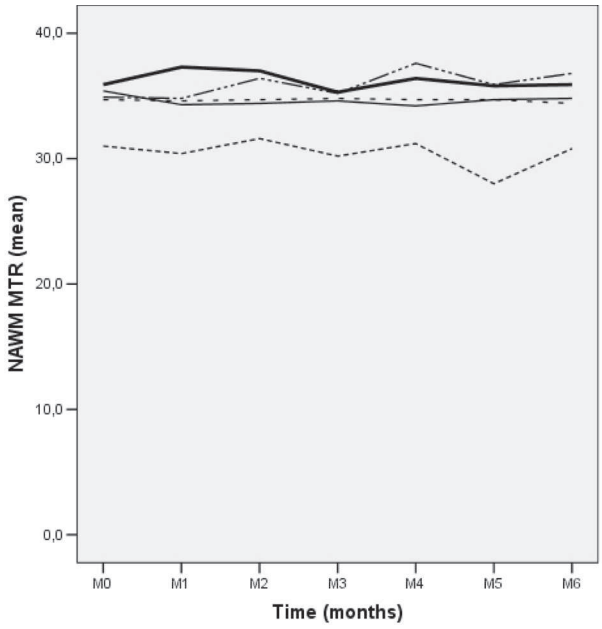

Figure 2 | Evolution of average Normal appearing white matter (NAWM) MTR. NAWM MTR in relapsing remitting MS patients from the sites: Munchen (bold), Graz (striped), Amsterdam (line), Milan (dotted) and Basel (dot-line) obtained on 7 monthly occasions.

Table 2 shows the empirical average lesion MTR for each consecutive timepoint before and after lesion enhancement. From a total of 457 newly enhancing lesions which developed during the 7 month follow-up, 349 lesions were assessed for MTR evolution. Since lesions were assessed for a maximum of seven timepoints, lesions cannot be followed from 6 months pre- to 5 months post-enhancement resulting in a gradual decline of the number of ROIs before and after timepoint 0. From all the MTR images acquired, approximately 90% were usable for analysis. Consistent with previous observations [3,4], MTR in the ROIs preceding future enhancement gradually decreased before lesion occurrence. At the time of enhancement, average lesion MTR dropped considerably to 25.4% (SD=6.03) and in the months after cessation of enhancement, the average lesion MTR gradually increased. The MTR evolution described above is visualized in Figure 3, together with the mean lesional MTR evolution of individual patients, illustrating the heterogeneity in lesional MTR evolution between patients.

Table 2 | Evolution of average lesional MTR. Observed average lesion MTR of relapsing remitting MS patients obtained on monthly MRIs of white matter areas before their future enhancement (month -6 to -1), during enhancement (month 0) and after enhancement (month 1 to 5).

Follow up (Months)	Mean lesion MTR (SD)	Number of ROIs
-6	31.6 (4.1)	54
-5	30.5 (4.9)	115
-4	31.1 (4.9)	171
-3	31.2 (4.1)	201
-2	30.9 (5.0)	243
-1	30.6 (4.9)	314
0	25.2 (5.4)	349
1	25.3 (5.2)	230
2	26.5 (5.5)	180
3	26.1 (5.6)	140
4	27.0 (5.6)	101
5	25.0 (6.5)	43

MTR = Magnetization Transfer Ratio
SD = Standard Deviation
ROI = Region Of Interest

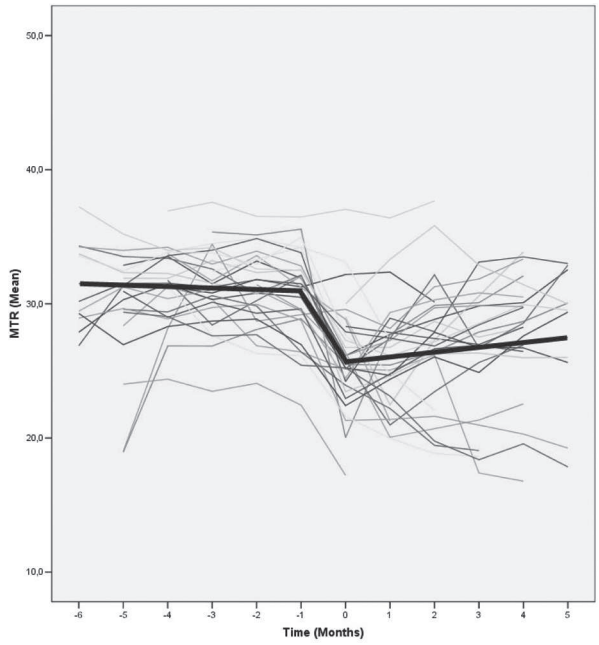

Figure 3 | Evolution of average lesional MTR. Evolution of lesional MTR surrounding the time of enhancement in individual patients (lines), and lesion MTR averaged over all patients based on a multilevel model (bold line).

Power calculations

The results of the multilevel estimates of the variance components introduced by each of the variance sources (lesion, patient, centre) are shown in Table 3. The variance for centres ($\sigma 2_{centres}$=0.73), patients ($\sigma 2_{patients}$=2.71), lesions ($\sigma 2_{lesions}$=12.58) and the residual variance ($\sigma 2_{residual}$=6.71), illustrate that the variation introduced by centres is considerably less than the variance introduced by patients or lesions. The difference in MTR between the last MTR measurement post enhancement and the first MTR measurement pre-enhancement was 4.0 %, and was assumed the 100% increase of lesional MTR post-enhancement, serving as the base for the simulated effects.

The estimated sample sizes are shown in **Table 4** for a varying magnitude in patient-treatment interaction and varying mean number of lesions per patient. Based

on the range of variance estimations above, we assumed the variance introduced by the difference in treatment-response between patients to be medium sized and not to surpass the variance of patients, and subsequently show sample sizes for no interaction (patient-treatment interaction of 0), mild interaction (patient-treatment interaction of 1.0) and moderate interaction (patient-treatment interaction of 2.5). As expected, when an interaction is present, the sample size increases in line with the magnitude of the interaction. A higher mean number of lesions per patient has a favourable effect on the estimated number of required patients. The calculations show that for a treatment effect of 30% increase in MTR post-enhancement in patients with a mean number of 6 lesions and the presence of a mild variation in treatment response between patients approximately 136 patients per trial are needed, whereas for a 50% increase in MTR approximately 48 subjects per trial are required.

Table 3 | Results of multilevel analysis of lesional MTR measurements. The fixed effects are the slope of MTR decrease and increase pre- and post- enhancement. The random effects describe the variance introduced by centres, patients, lesions and the residual variance.

Effect		Estimate	s.e.
Fixed	Intercept	28.3	0.59
	Pre-post*	-2.6	0.29
	Slope pre enhancement	-0.10	0.05
	Slope post enhancement	0.36	0.07
Random*	Centres	0.73	1.03
	Patients within centres	2.71	1.26
	Lesions within patients	12.59	1.11
	Centres x pre-post	0	
	Patients x pre-post	1.61	0.57
	Lesions x pre-post	2.28	0.31
	residual	6.71	0.25

*covariances not shown; prepost: -1=pre, +1=post

Table 4 | Sample size estimates. Sample size estimates required to significantly detect effects of treatments enhancing the recovery of newly formed lesion MTR post enhancement in a placebo controlled clinical trial of relapsing remitting MS patients. The variance in treatment response between patients is assumed 0, mild (1) and moderate (2.5). Estimates are the total number of patients for the complete trial, based on a cohort of 100% active patients, and 0% dropouts.

variance in treatment response between patients	Mean number of lesions per patient	Treatment effect size (% increase in MTR post enhancement)				
		30%	40%	50%	60%	70%
None	2	144	84	52	36	28
	4	72	44	28	20	16
	6	48	28	20	12	12
	8	36	24	16	12	8
Mild	2	232	132	84	60	44
	4	160	92	60	40	32
	6	136	76	48	36	28
	8	124	72	44	32	24
Moderate	2	360	204	132	92	68
	4	288	164	104	72	56
	6	264	148	96	68	52
	8	252	144	92	64	48

MTR = Magnetization transfer ratio

DISCUSSION

The shift in interest from anti-inflammatory towards neuroprotective and neurorepairative treatments in MS clinical trials has increased the demand for outcome measures with a higher pathological specificity for neuronal damage and repair. In this respect, magnetization transfer imaging, with its specificity for demyelination and remyelination, is a promising addition for conventional scanning protocols, being a quantitative and reproducible technique, allowing both global and focal assessment of disease progression and being implementable on most MRI scanner types. The present study shows that multicentre clinical trials using lesional MTR recovery as primary outcome for measuring changes over time in newly formed lesions require feasible sample sizes which are applicable in a phase II setting.

Inter-centre variations for MTR measurements are a familiar phenomenon in multicentre trials due to the dependency of the acquired MTR data on the pulse sequence parameters of the scanners applied [12,13]. Although our data contained MTR measurements from five centres and were obtained without an MTI standardization protocol, the multilevel analysis showed that the variance in MTR introduced by multiple centres constituted only a minor part of the total variance; considerably greater variance in MTR was caused by differences between patients and between lesions.

To measure the MTR of new enhancing lesions pre and post enhancement in the context of a clinical trial, ideally one wants to synchronize the start of a treatment with the start of enhancement and in the present analyses therefore, treatment is assumed to be effective from the time of enhancement and of no influence beforehand. In practice this implies that only a proportion of lesions, *e.g.* appearing in one or two months at the start of treatment can be assessed for MTR evolution. After a run-in period of 6 months for example, all enhancing lesions in the subsequent one or two months are followed and scanned for another 4 months, yielding a complete trial duration of 1 year. In this way, the mean number of enhancing lesions picked up during the trial will prove lower compared to a design where time of enhancement and start of treatment are not standardized, and the estimated sample sizes for the lower mean number of lesions should be followed.

Some considerations should be taken into account when the current sample sizes are applied in practice. First, one should realize that the current calculations assume solely active patients to participate in the trial. When a percentage of inactive patients or dropouts is expected in the design-phase of a future trial, the given sample size should be recalculated accordingly (eq. 2). Second, the impact of the mean number of lesions on the required sample sizes in our analyses signifies the need for recruitment of active patients in future trials, *e.g.* selecting patients with one or more enhancing lesions on the baseline scan, to yield sufficient data points, stabilize the variability and increase the study power. Third, the lesional MTR measurements in this study were acquired by co-registration of the lesion ROIs with the magnetization transfer imaging scans. While this is a robust procedure, the interpolation due to the transformations in the registration could have introduced unwanted averaging between neighbouring pixels, appearing as noise in the ROI mean MTR values and thereby increasing the variance of our measurements. Nevertheless, the power calculations based on this data still yielded applicable sample sizes. Application of more sophisticated (*e.g.* 3D

scanning) acquisition protocols and more accurate ROI analysis procedures will likely reduce the measurement variance and improve the sample sizes even further. Another consideration is our assumption of a 100% treatment effect yielding an MTR value similar to the starting MTR value pre-enhancement. A previous post-mortem study [14] showed that fully remyelinated lesions returned a significantly lower MTR signal than the MTR of the corresponding NAWM. Whilst the MTR of the NAWM in our data is still considerably higher than the starting MTR value pre-enhancement (which we assume the maximal attainable MTR value), one should be aware that the measurable effect-sizes based on these findings become smaller, and require larger sample size. This implies that for the current sample size estimates a simulated treatment effect of *e.g.* 50% is already of considerable size.

To our knowledge, this is the first study reporting sample size estimates for clinical trials using lesional MTR as outcome measure. Compared to the sample size estimates for trials using alternative measures for neuroprotection and repair, the use of lesional MTR recovery appears to be a favourable measure in terms of study power. Anderson *et al.* [15] showed that trials using the rate of cerebral atrophy (acquired using SIENA) as primary outcome measure required approximately 190 patients per arm to detect a 30% decrease in atrophy rate over a 1 year follow-up period, and 111 patients per arm over 3 years of follow-up. Another recent study showed sample size estimates for trials using the number of Persistent Black Holes (PBH) as primary outcome measure [16]. Here, approximately 200 patients per treatment arm are needed to detect a 50% reduction in mean number of PBHs, while detection of smaller effect sizes would require considerably more patients. Not only do both outcome measures require a larger sample size to detect significant treatment effects, acquisition of data takes considerably more time (atrophy) or requires a potent drug to determine a treatment efficacy with acceptable sample size (PBH). An important advantage of using atrophy as outcome measure however, is the requirement of considerably less MRIs (annual, rather than monthly).

The evolution of enhancing lesions in terms of hypo- and iso-intensinty on the corresponding unenhanced T1w image has been shown to be a patient bound phenomenon, with lesion evolution being uniform within patients [17]. When assessed by magnetization transfer imaging, the heterogeneity of lesion evolution is more prominent: in addition to differences *between* patients, MTR variablity is mainly driven by differences between lesions *within* patients. This reflects the more

accurate measurement of a quantitative continuous outcome measure (lesional MTR). compared to a categorical outcome measure (iso or hypo- intensity assessment).

A limitation of our study is the inability to estimate the aforementioned variances introduced by the differences in response to treatment between patients or between centres, since their estimation requires lesional MTR data with an actual treatment effect, which is presently unavailable. While the assumption of equal responses to treatment between centres is justifiable given that MS patients in centres are randomly sampled from the overall MS population, the assumption of equal treatment responses between patients is less easily justified, knowing that the effect of current approved therapies may vary between MS patients. Although the current calculations based on the presence of patient-treatment interaction reflect the required number of patients for future studies more realistically, we underline the importance of exploring the possibility of interaction between treatment and patients in lesional MTR trial data.

The sample size calculations are based on the estimated individual level variance-components of an active cohort of RRMS patients originally selected for participation in a clinical trial, with common baseline demographic, clinical and MRI characteristics, serving as a representative sample of MS patients for modelling the evolution of lesional MTR.

Still, with only the present cohort at our disposal, we were not able to assess the consistency of the individual variance component estimations in different RRMS datasets, which is a desirable issue for future research for validating the present sample size calculations. The exclusion of small lesions not distinguishable from non-lesion brain tissue in this study could have slightly biased the results toward a more severe lesional MTR evolution, and repetition of the analyses with data from all visible lesions acquired with more sophisticated future techniques might prove interesting.

A promising development in this regard has recently been shown by Chen *et al*. [18] who, instead of mean lesion MTR, have developed a method to monitor the evolution of MTR changes of individual lesion voxels. By quantifying the evolution over time of those voxels undergoing significant decreases and increases in MTR consistent with demyelination and remyelination respectively, this method is potentially more sensitive for picking up effects of experimental treatments, thereby enhancing study power, although interpolation issues including those discussed above should also be expected to influence this method.

In conclusion, the sample sizes calculated in this study show that lesional MTR recovery is a feasible primary outcome measure for future multicentre phase II trials testing the efficacy of remyelinating treatments in active RRMS patients.

ACKNOWLEDGEMENTS

We are grateful to Bayer-Schering Pharma AG, Berlin, Germany and Rentschler Biotechnology GmbH, Laupheim, Germany for allowing us to use the data of the oral Interferon study. G.K. was supported by an educational grant by Bayer Schering Pharma.

FUNDING

The MS Center Amsterdam is supported by the Dutch MS Research Foundation (Grant. no. 05-358c).

REFERENCES

1. **Brück W, Bitsch A, Kolenda H, Brück Y, Stiefel M, Lassmann H.** Inflammatory central nervous system demyelination: correlation of magnetic resonance imaging findings with lesion pathology. *Ann Neurol*. 1997; 42(5): 783-793.
2. **Schmierer K, Scaravilli F, Altmann DR, Barker GJ, Miller DH.** Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Ann Neurol*. 2004; 56(3): 407-415.
3. **Filippi M, Rocca MA, Martino G, Horsfield MA, Comi G**. Magnetization Transfer Changes in the normal appearing white matter precede the appearance of enhancing lesions in patients with multiple sclerosis. *Ann Neurol*. 1998; 43: 809-814.
4. **Fazekas F, Ropele S, Enzinger C, Seifert T, Strasser-Fuchs S.** Quantitative magnetization transfer imaging of pre-lesional white-matter changes in multiple sclerosis. *Mult Scler.* 2002; 8(6): 479-484.
5. **van Waesberghe JH, van Walderveen MA, Castelijns JA, et al.** Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin-echo and magnetization transfer MR. *Am J Neuroradiol*. 1998; 19(4): 675-683.
6. **Filippi M, Rocca MA, Sormani MP, Pereira C, Comi G.** Short-term evolution of individual enhancing MS lesions studied with magnetization transfer imaging. *Magn Reson Imaging*. 1999; 17(7):979-984.
7. **Agosta F, Rovaris M, Pagani E, Sormani MP, Comi G, Filippi M.** Magnetization transfer MRI metrics predict the accumulation of disability 8 years later in patients with multiple sclerosis. *Brain* 2006; 129(10): 2620-7.
8. **Kita M, Goodkin DE, Bacchetti P, Waubant E, Nelson SJ, Majumdar S.** Magnetization transfer ratio in new MS lesions before and during therapy with IFNbeta-1a. *Neurology* 2000; 54(9):1741-1745.
9. **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.
10. **Polman C, Barkhof F, Kappos L, Pozzilli C, Sandbrink R, Dahlke F** et al. and the European Oral Interferon Beta-1a in Relapsing-Remitting MS Study Group. Oral interferon beta-1a in relapsing-remitting multiple sclerosis: a double-blind randomized study. *Mult Scler.* 2003; 9(4):342-348.
11. **Snijders TAB**: Sampling. In: **Leyland AH, Goldstein H.** *Multilevel modelling of health statistics*, Wiley, New York 2001.
12. **Tofts PS, Steens SC, Cercignani M, et al.** Sources of variation in multi-centre brain MTR histogram studies: body-coil transmission eliminates inter-centre differences. *MAGMA*. 2006 19(4): 209-222.
13. **Berry I, Barker GJ, Barkhof F, et al.** A multicenter measurement of magnetization transfer ratio in normal white matter. *J Magn Reson Imaging*. 1999; 9(3):441-446.
14. **Barkhof F, Bruck W, De Groot CJA, et al.** Remyelinated lesions in multiple sclerosis; magnetic resonance appearance. *Arch Neurol*. 2003; 60:1073-1081.

15. **Anderson VM, Bartlett JW, Fox NC, Fisniku L, Miller DH.** Detecting treatment effects on brain atrophy in relapsing remitting multiple sclerosis: sample size estimates. *J Neurol*. 2007; 254(11): 1588-1594.

16. **van den Elskamp IJ, Lembcke J, Dattola V, et al.** Persistent T1 hypointensity as an MRI marker for treatment efficacy in multiple sclerosis. *Mult Scler.* 2008; 14(6):764-769.

17. **Minneboo A, Uitdehaag BM, Ader HJ, Barkhof F, Polman CH, Castelijns JA.** Patterns of enhancing lesion evolution in multiple sclerosis are uniform within patients. *Neurology* 2005; 12; 65(1): 56-61.

18. **Chen JT, Collins DL, Atkins HL, Freedman MS, Arnold DL.** Magnetization transfer ratio evolution with demyelination and remyelination in multiple sclerosis lesions**.** *Ann Neurol*. 2008;63(2): 254-262.

Appendix

Sample size calculations

Suppose a parameter d of a (multilevel) normal linear (regression) model measures the effect of interest. Then, the relation between the variance of the estimator, denoted by var($\hat{\delta}$), and power is approximately

$$(1) \qquad \frac{\delta^2}{var(\hat{\delta})} \approx (z_{1-\alpha/2} + z_{1-\beta})^2,$$

where a is the two-sided significance level, 1-ß the power and z(.) are percentile points of the standard normal distribution (Snijders, 2001).

It should be noted that for an intended placebo controlled trial the randomization is at the patient level. Furthermore, respecting the multilevel structure of the data (MTR measurements nested within lesions, lesions within patients and patients within centres), Table A1 shows the sources of variation and the model terms that can be distinguished for the maximal model using a factor relationship diagram (Bergerud, 1996).

Table A1.

Source	Model term	F/R*	Parameter
Intercept	γ_0	F	γ_0
Centres	f_{i0}	R	σ^2_{f0}
Treatment (X)	$\gamma_1 X_{ij}$	F	γ_1
Centres X Treatment	$f_{i1} X_{ij}$	R	σ^2_{f1}
Patients	$v_{ij\,0}$	R	σ^2_{v0}
Lesions	u_{ijk}	R	σ^2_u
Prepost (Z)	$\gamma_2 Z_{ijkm}$	F	γ_2
Centres X Prepost	$f_{i2} Z_{ijkm}$	R	σ^2_{f2}
Treatment X Prepost	$\gamma_3 X_{ij} Z_{ijkm}$	F	γ_3
Centres X Treatment X Prepost	$f_{i3} X_j Z_{ijkm}$	R	σ^2_{f3}
Patients X Prepost	$v_{ij1} Z_{ijkm}$	R	σ^2_{f3}
Residual	e_{ijkm}	R	σ^2_e

* F = fixed, R = random

** Treatment = 1: intervention, Treatment = -1: placebo

*** Prepost = 1: post measurement, Prepost = -1: baseline measurement

In this design, the effect of primary interest d is the difference between the means of the measurements of the intervention and the placebo group of patients, adjusted for differences at baseline, i.e. essentially the treatment ´ time interaction. It can be easily shown that $\delta=4\gamma_3$.

In a balanced design, an unbiased estimator of d is

$\hat{\delta}=\overline{Y}$(intervention, post measurement)$-\overline{Y}$(placebo, post measurement) $-$
[$=\overline{Y}$(internvetion, baseline measurement) $-$ $\overline{Y}$(placebo, baseline measurement),

where $\overline{Y}$ denotes the mean of all MTR measurements under the specified condition. Following the same reasoning as in Snijders (2001), the variance can be shown to be equal to

$$(2) \qquad \text{var}(\hat{\delta})=\text{var}(4\hat{\gamma}_3)=16\,\frac{2\eta_2\eta_3\sigma^2_{f3}+2\eta_2\sigma^2_{v1}+\sigma^2_e}{2\eta_2\eta_3\eta_4},$$

where η_4, η_3 and η_2 and η_2 are the number of centres, the number of patients per centre and the number of lesions per patient, respectively. Note that $2\eta_2\eta_3\eta_4$ is the total number of measurements, and that η_3 should be *even*, because half of the patients per centre will be assigned to the intervention, and the other half to the placebo condition.

Substituting (2) in (1) and rewriting yields

$$(3) \qquad \eta_2\eta_3\eta_4=8\,\frac{\approx(z_{1-\alpha/2}+z_{1-\beta})^2}{\delta^2}\,(2\eta_2\eta_3\sigma^2_{f3}+2\eta_2\sigma^2_{v1}+\sigma^2_e)$$

REFERENCES

- **Bergerud WA** (1996). Displaying factor relationships in experiments. *The American Statistician 1996*; 50(3): 228-233.
- **Snijders, TAB** (2001). Sampling. In: AH Leyland and H Goldstein. *Multilevel Modelling of Health Statistics*. New York: Wiley. Chapter 11: 159-174.