

# Exploratory treatment trials in multiple sclerosis using MRI: sample size calculations for relapsing-remitting and secondary progressive subgroups using placebo controlled parallel groups

N Tubridy, H J Ader, F Barkhof, A J Thompson, D H Miller

## Abstract

**Objectives**—Serial brain MRI is widely used in pilot studies of new agents to monitor treatment efficacy in relapsing-remitting (RR) and secondary progressive (SP) multiple sclerosis (MS). For pilot trials, sample size calculations for the RR subgroup are based on the data from small numbers of patients and separate calculations for the SP subgroup have not been performed. The present study considers these issues.

**Methods**—The sample size calculations were based on data from six months of monthly T2 weighted and gadolinium enhanced MRI in 31 RR and 28 SP untreated patients undergoing natural history studies or in the placebo arm of a therapeutic trial. The calculations were for a placebo controlled, parallel groups design lasting six months. The sample sizes were based on bootstrap analysis with an 80% likelihood of showing a given treatment effect.

**Results**—With a single pretreatment scan, demonstration of a 70% reduction in newly active lesions required 2×30 RR and 2×50 SP patients. With an extra run-in scan one month before treatment, the sample sizes were 2×20 for RR and 2×30 for SP patients.

**Conclusions**—The sample sizes required for RR patients were comparable with previous smaller studies. Larger sample sizes were needed for the SP group, but the extra run in scan resulted in a reduction in both groups. The larger sample sizes in the SPMS group were probably due to the combination of a higher proportion of patients with low MRI activity ( $\leq 2$  active MRI lesions in 50% of SP and 32% RR patients), as well as a few patients who displayed extremely high activity, thus increasing interpatient variability. These data should be considered in planning pilot MRI outcome trials.

(*J Neurol Neurosurg Psychiatry* 1998;64:50-55)

Keywords: multiple sclerosis; power calculations; treatment trials

Because of the highly variable course of multiple sclerosis (MS) and the unsatisfactory

nature of the commonly used measures of outcome, there are inherent problems in designing treatment trials with clinical end points such as relapse rate or change in disability.<sup>1-3</sup> Clinical end points are definitive but are logistically difficult to reach. Therefore, surrogate markers that give earlier answers are much sought after.

Magnetic resonance imaging (MRI) is currently the most plausible surrogate marker for monitoring disease activity in treatment trials.<sup>2,4</sup> In relapsing-remitting (RR) and secondary progressive (SP) MS, monthly T2 weighted and/or gadolinium-enhanced brain MRI detect on average about five to 10 new lesions for every clinical relapse.<sup>5-11</sup> Based on natural history data with small numbers of patients, sample sizes for pilot MRI studies have been calculated.<sup>12-14</sup> The calculations have been based on data from patients with RRMS alone or on RR combined with SP patients. A separate set of power calculations for patients with SPMS alone has not been performed before. The present study is based on a larger cohort and reports calculations for the RRMS and SPMS groups separately.

## Methods

The data from patients involved in five previous studies were reviewed.<sup>7,8,10,11,15</sup> Serial scanning had been performed either as part of a natural history study,<sup>7,8,10,11</sup> or in patients in the placebo arm of a therapeutic trial.<sup>15</sup> During the study period, none of the patients had received disease modifying therapy, apart from short courses of corticosteroids for acute relapses. Each patient had initial T2 weighted and gadolinium enhanced brain MRI and then monthly follow up scans for six months.

Two groups of patients were defined: (1) RRMS (n=31). This group was defined as clinically definite MS<sup>16</sup> with clear episodes of acute neurological dysfunction lasting more than 24-48 hours with full or partial recovery and a stable clinical state between attacks.<sup>17</sup> The clinical and MRI data of these patients were drawn from three previously published sources.<sup>8,10,15</sup> (2) SPMS (n=28). This was defined as clinically definite MS<sup>16</sup> with the development of a gradual increase in disability, with or without superimposed relapses, for at least six months, after an initial relapsing-remitting course.<sup>17</sup> The data were also drawn from three sources.<sup>7,11,15</sup>

NMR Research Unit,  
Institute of Neurology,  
Queen Square,  
London, UK  
N Tubridy  
A J Thompson  
D H Miller

Department of  
Biostatistics  
H J Ader

MR Centre for  
Multiple Sclerosis  
Research, Free  
University Hospital,  
Postbus 7057, 1007 MB,  
Amsterdam, The  
Netherlands  
F Barkhof

Correspondence to:  
Professor DH Miller, NMR  
Unit, Institute of Neurology,  
Queen Square, London  
WC1N 3BG, UK. Telephone  
0044 171 837 3611; fax  
0044 171 278 5616.

Received 3 April 1997 and in  
revised form 23 July 1997

Accepted 5 August 1997

Table 1 Clinical data for RRMS and SPMS patients

Clinical data	RRMS (n=31)	SPMS (n=28)
Mean (range) age (y)	30.1 (22-45)	41.4 (24-53)
Sex (M:F)	7:24	16:11
Mean (range) disease duration (y)	3.7 (1-8)	10.6 (2-26)
Mean (range) EDSS (at entry)	2.8 (1-6)	5.5 (3.5-8)

RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis; EDSS = expanded disability status scale.

Table 2 Number of enhancing lesions on baseline MRI and newly active lesions on six months of follow up scans in RRMS patient

Patient No	Baseline scan	Months of follow up						AII	AIII
		1	2	3	4	5	6		
1	4	0	0	0	0	0	0	0	0
2	7	0	4	3	5	10	17	83	39
3	0	0	0	2	0	13	6	50	21
4	4	2	1	2	7	4	4	100	20
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	1	1	0	0	1	0	50	3
8	0	0	0	0	0	0	0	0	0
9	1	4	2	2	0	0	1	67	9
10	1	0	1	1	3	1	0	67	6
11	5	6	4	4	7	9	7	100	37
12	0	0	0	0	0	0	0	0	0
13	0	0	0	0	1	0	0	17	1
14	4	2	1	0	0	0	0	33	3
15	2	2	2	1	2	4	1	100	12
16	0	4	1	4	0	0	0	50	9
17	1	1	0	0	0	1	0	33	2
18	0	0	0	0	0	0	0	0	0
19	2	2	0	0	0	0	0	17	2
20	0	0	2	2	2	0	0	50	6
21	0	0	5	1	1	1	0	67	8
22	0	1	0	7	5	11	12	83	36
23	0	0	0	1	2	0	0	33	3
24	1	1	1	0	1	0	1	67	4
25	5	4	0	1	0	2	0	50	7
26	0	2	0	0	0	0	0	17	2
27	3	11	5	3	2	1	0	83	22
28	0	1	1	1	0	0	0	50	3
29	0	3	0	0	1	1	2	67	7
30	3	2	0	0	0	1	1	50	4
31	1	1	1	1	3	0	1	83	7

AII = the percentage of scans (per patient) on which newly active lesions were seen.  
AIII = the total number of active lesions seen during the study period.

Table 3 Number of enhancing lesions on baseline MRI and newly active lesions on six months of follow up scans in patients with SPMS

Patient No	Baseline scan	Months of follow up						AII	AIII
		1	2	3	4	5	6		
1	0	0	1	1	0	0	0	33	2
2	0	0	0	0	0	0	1	17	1
3	1	0	1	5	3	10	12	83	31
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	1	1	0	0	0	33	2
8	0	3	1	2	1	2	1	100	10
9	1	0	1	0	0	0	0	17	1
10	0	0	0	0	0	1	3	33	4
11	0	0	0	0	0	0	0	0	0
12	0	0	1	0	2	4	2	67	9
13	0	0	0	1	2	4	2	67	9
14	16	13	13	5	1	0	7	83	39
15	3	2	4	11	3	9	2	100	31
16	6	2	3	6	3	4	1	100	19
17	5	6	5	2	3	2	2	100	20
18	10	15	9	5	7	9	21	100	66
19	0	0	0	3	0	2	2	50	7
20	1	0	0	0	0	0	1	17	1
21	0	3	36	4	17	10	15	100	85
22	1	0	4	0	0	1	0	33	5
23	0	0	0	0	1	1	0	33	2
24	0	0	0	0	0	0	0	0	0
25	1	0	0	0	0	0	0	0	0
26	0	1	0	0	0	0	0	17	1
27	0	0	0	0	0	0	1	17	1
28	0	1	0	0	3	0	1	50	5

AII = the percentage of scans (per patient) on which newly active lesions were seen.  
AIII = the total number of active lesions seen during the study period.

Table 1 shows the clinical data for the patients in each group. The patients had a wide range of duration of disease and disability but overall had clinical characteristics typical of both disease subgroups. The patients, as initially selected, were representative of MS patients seen at a tertiary referral centre and were also representative of patients who would be considered for inclusion in treatment trials. Although the data are drawn from several studies, scans were performed using comparable protocols. Thus all the patients had a T2 weighted (TR 2000-2755 ms, TE 60-120 ms) scan. A T1 weighted image (TR 450-575 ms, TE 13-40 ms) was obtained after 0.1 mmol/kg gadolinium -DTPA was given intravenously.

On the pretreatment scan, the number of gadolinium enhancing lesions was counted. On the follow up scans newly active lesions were counted. These were defined as new enhancing lesions (>90% of all new active lesions), new non-enhancing lesions, or new enlarging but non-enhancing lesions on the T2 weighted images. Table 2 and table 3 show the serial scan data tabulated for both patient groups.

Three MRI outcomes were evaluated: AI—the number of patients showing new active lesions at any time during the study period; AII—the proportion of scans showing newly active lesions during the study; AIII—the number of newly active lesions seen over the whole study period. AI data are not considered further as they show much poorer statistical power.<sup>13</sup>

Power estimates were then calculated using a “bootstrap” method of computerised sampling and trial simulation as previously described.<sup>18</sup> In this procedure, for every study under consideration 1000 cases are drawn randomly with replacement (each patient may be repeatedly drawn) from the original data sets. A theoretical distribution is used to simulate a treatment effect. In the case of a homogenous population being assumed, use is made of a Bernoulli trial (“coin flipping” as it is referred to in the article). When a heterogenous (more variable) patient response to a proposed treatment is more plausible, a  $\beta$  distribution is used which accounts for differences between patients, but still allows for a mean probability of response to be calculated. To compute the resulting power of a given sample size at a certain treatment efficacy, Wilcoxon’s test statistic is used to compare the (simulated) patient groups. The corresponding probability is used as a power estimate.

To arrive at confidence intervals for the power estimates given in the tables in this report, the entire procedure could be repeated, for instance, 100 times. To compute such confidence intervals the procedure should be repeated a sufficient number of times and the resulting power estimates analysed anew. As all bootstrap sample sizes that are calculated will be based on the same underlying patient data, appropriate corrections on the final variance should be made.

A total of eight sets of power calculations were computed—four for each of the two patient groups. Power calculations were made

Table 4 Power calculations using AIII response variable for RRMS patients (n=31), without a run-in scan (homogeneous response only)

Sample size	Efficacy (%)	Months of follow up			
		1	2	4	6
2×20	70	34	47	59	61
	80	56	68	81	83
2×30	60	35	41	55	63
	70	53	64	77	81
2×40	80	71	86	92	96
	60	43	52	70	72
2×50	70	64	77	86	89
	80	82	95	99	99
2×75	60	51	64	77	83
	70	72	84	93	95
2×100	80	93	97	99	100
	60	70	80	90	92
2×100	70	91	96	99	99
	80	98	100	100	100
2×100	60	83	89	97	98
	70	96	98	100	100
2×100	80	100	100	100	100

Table 5 Power calculations using AIII response variable for RRMS patients (n=31), with a run-in scan (homogenous response)

Sample size	Efficacy (%)	Months of follow up			
		1	2	4	6
2×20	70	28	58	65	85
	80	43	77	86	97
2×30	60	36	56	66	83
	70	43	75	84	95
2×40	80	55	91	96	100
	60	41	68	78	91
2×50	70	52	88	94	99
	80	67	97	99	100
2×75	60	48	83	85	96
	70	65	93	97	100
2×100	80	76	98	100	100
	60	69	94	96	99
2×100	70	80	99	100	100
	80	91	100	100	100
2×100	60	78	98	99	100
	70	91	100	100	100
2×100	80	97	100	100	100

Table 6 Power calculations for SPMS patients (n=28) using AIII response variable without an extra baseline correction scan (homogenous response)

Sample size	Efficacy (%)	Months of follow up			
		1	2	4	6
2×20	70	23	28	35	40
	80	39	43	52	59
2×30	60	21	25	30	37
	70	37	38	49	55
2×40	80	55	64	71	76
	60	29	32	43	44
2×50	70	43	52	63	70
	80	68	74	84	88
2×75	60	35	37	48	56
	70	56	58	73	78
2×100	80	78	84	91	95
	60	47	55	67	72
2×100	70	70	78	86	91
	80	93	94	98	99
2×100	60	62	67	76	85
	70	84	89	94	97
2×100	80	97	98	100	100

both with and without the use of an additional run-in scan obtained one month before the start of treatment. Using this extra scan, the new active lesions between month-1 and month 0 (the pretreatment scan) were subtracted from the number of new active lesions during the study period. Calculations were made for both a homogenous (little between patient variability), and for a heterogenous (where patient response is highly variable) treatment response.

The efficacy of an experimental treatment was expressed as a percentage and represents the reduction in newly active lesions (AIII) seen when a patient has received a treatment compared with those given placebo. The sample size was calculated for a 60%, 70%, and 80% level of efficacy. Only a parallel groups study design was simulated. It was assumed that the putative treatment takes immediate effect and that its efficacy does not alter during follow up.

## Results

In the RRMS group (n=31), on the initial scan 15 (48.4%) patients had enhancing lesions. During follow up, six patients showed no newly active lesions during the study period. The total number of scans with newly active lesions was 88 (47.3%). The mean number of newly active lesions/patient was 8.9 (median 4 (SE 2), range=0–39).

In the RRMS group, without the extra run-in scan, the demonstration of a 70% AIII efficacy with a power of 80 required 2×30 patients to be studied for six months or 2×40 patients for four months (table 4). When the extra run-in scan was added, 70% efficacy was demonstrated if 2×20 patients were studied for six months or 2×30 patients for four months (table 5).

In the SPMS group (n=28), on the initial scan 10 (35.7%) had enhancing lesions. During follow up, six patients showed no newly active lesions in the six months of the study. The total number of scans with newly active lesions was 75 (44.6%). The mean number of newly active lesions/patient was 12.2 (median 2.0 (SE 3.9), range=0–85).

In the SPMS group without the run-in scan, demonstration of a 70% AIII efficacy with a power of about 80 required 2×50 patients to be studied for six months (the actual power calculated was 78 for a homogeneous and 83 for heterogeneous response) or 2×75 for four months (table 6). When the extra run-in scan was added, 70% efficacy was demonstrated in 2×30 patients in four or six months, but not in 2×20 patients studied for six months (table 7): thus the SPMS sample sizes come closer to those seen in the RR group (figure).

Three groups of patients with different levels of new lesion activity were defined:

(1) *Low activity*—0–2 new active lesions during the six months of follow up:

(2) *Moderate activity*—3–30 new active lesions.

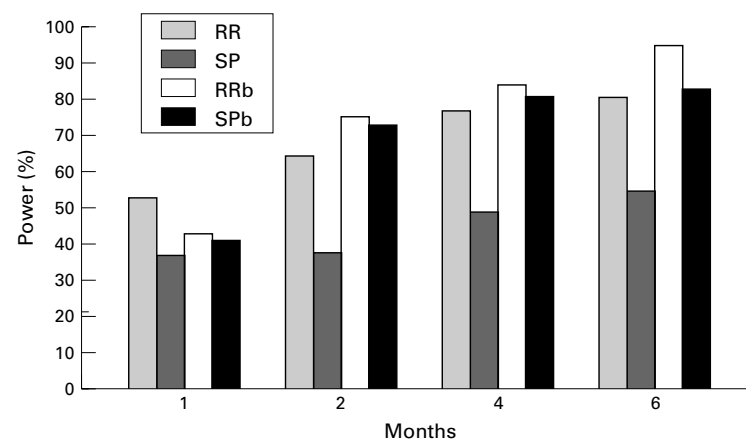
(3) *High activity*—greater than 30 new active lesions.

A higher proportion of SPMS patients exhibited low activity (14/28 (50%) v 10/31(32%) of RR patients), whereas a greater number of RR patients had moderate activity (18/31 (58%) v 9/28 (32%) of SP patients). High activity was seen in 18% of SP and 10% of RR patients. The two patients with the highest MRI activity (66 and 85 new active lesions) had SPMS (table 8).

The sample size calculations for the two groups of patients were tabulated for the AIII response (tables 4–7). The calculations for a

Table 7 Power calculations using AIII response variable for SPMS patients (n=28) with a run-in scan (homogenous response)

Sample size	Efficacy (%)	Months of follow up			
		1	2	4	6
2×20	70	29	53	66	67
	80	42	76	83	90
2×30	60	34	56	63	57
	70	41	73	81	83
2×40	80	53	88	97	98
	60	36	66	76	73
2×50	70	54	86	93	92
	80	67	94	99	100
2×75	60	50	78	83	83
	70	64	91	95	97
2×100	80	78	99	100	100
	60	64	91	94	93
2×100	70	78	98	100	100
	80	92	100	100	100
2×100	60	78	97	98	99
	70	90	100	100	100
2×100	80	96	100	100	100



Power calculations for subgroups with sample size 2×30 and 70% efficacy.

Table 8 New MRI activity during months 0-6 (n (%))

Subgroup	Low activity	Moderate activity	High activity
RRMS	10 (32)	18 (58)	3 (10)
SPMS	14 (50)	9 (32)	5 (18)

Low activity = 0-2 new active lesions; moderate activity = 3-30 new active lesions; high activity = > 30 new active lesions.

homogeneous or heterogeneous response were generally very similar: only the homogeneous response data are presented. Slightly smaller sample sizes were generally required for AIII than for AII data, in both RR and SP subgroups (AII data not shown).

### Discussion

At present, the generally modest correlations between MRI findings and clinical status in MS means that new therapies are ultimately judged in large phase III trials in which a clinical outcome is the primary measure.<sup>4</sup> Nevertheless, MRI is now widely used as the primary outcome measure of disease activity in exploratory Phase I/II trials in RRMS and SPMS, because the high frequency of asymptomatic disease activity detected by MRI in these subgroups makes it a powerful tool for small cohort and short duration studies.<sup>5-11</sup> Furthermore, some correlations do exist between the activity seen on serial T2 weighted and gadolinium enhanced scans and clinical measures of disease activity or progression in both

RR<sup>5 6 8-10 19</sup> and SP<sup>8 20</sup> subgroups, the most consistent being the higher frequency of enhancing lesions during clinical relapse. These correlations suggest that it is valid and clinically relevant to use MRI to determine efficacy in exploratory trials.

Some previous MRI studies of small cohorts have reported broadly similar mean rates of new lesion activity in RR and SP disease.<sup>5-11</sup> Pronounced between patient variability in the level of MRI activity has, however, been readily apparent in both groups.<sup>5-12 21</sup> In particular, higher rates of activity have been reported in RR and SP patients who continued to relapse, whereas lower rates were seen in RR patients in remission<sup>10</sup> and SP patients who continue to progress, but without superimposed relapses.<sup>11</sup> Such heterogeneous patterns in reports involving small cohorts led us to the present study of larger and separate RR and SP cohorts on which to perform sample size calculations for exploratory MRI outcome treatment trials.

The number of patients that need to be enrolled into a trial, the number of hospital visits required, and the number of scans that need to be performed, must all be established to design a trial which minimises the burden on patients, is cost efficient, and has a high likelihood of showing the anticipated treatment effect. In the last respect, trials that are predicted to yield a statistical power less than 80 are generally deemed unacceptable. We employed well documented statistical techniques.<sup>12-14</sup> In the RR subgroup we compared our results with other published studies, but have used a larger patient cohort. Our study also considered SPMS patients as a separate entity for the first time.

Our results showed significant differences between the two groups. When just the single pretreatment scan was obtained, substantially larger sample sizes were required for the SP group to show an equivalent therapeutic effect in a given period of time.

The addition of one extra run-in scan (but not more than one) before entry into an exploratory treatment trial was previously shown to reduce the sample sizes required.<sup>13</sup> We again saw this effect in the present analysis, in both of our patient subgroups. For SPMS patients the effect was especially pronounced, and using the run-in scan the differences in sample sizes needed to show the same treatment effect in the same period of time were smaller (table 7).

The beneficial effect of the addition of the extra run-in scan is that it reduces some of the between patient variation in MR activity. The appreciably larger sample sizes in the SP patients when such a scan was omitted suggests a greater between patient variability among these patients. There were indeed notable differences between RRMS and SPMS groups when the new lesion activity over the six months was divided into those patients who showed low, moderate, and high levels of new lesion activity. Half of the SPMS group had low levels of activity compared with only one third of RRMS patients. By contrast, most RR patients (58%) displayed moderate new lesion



activity. High levels of activity occurred in a small proportion in each group but if anything slightly more in the SPMS group, a few of whom were extremely active. The occurrence of more low activity scans in the SPMS patients will reduce the statistical power in this group as there will be fewer lesions to “treat”. The occurrence of more extreme highs in the same group will also reduce the power of the study as there is a greater degree of between patient variability. These factors probably account for the larger sample sizes calculated for SPMS than for RRMS patients in the present study.

The present study provides tables (tables 4–7) of the power estimates for a range of sample sizes at treatment efficacies of 60%, 70%, and 80%. The efficacy of an experimental treatment is, for example 70%, if an average of 70% fewer active lesions are seen when a treatment is given compared with placebo. Treatment efficacies of less than 60% can be calculated by adapting the computer program in such a way that the user can enter any desired efficacy and run it with any data set they wish (the data set on which this article is based is included as a demonstration set (table 2 and table 3)). However, treatment efficacies lower than 60% may need to be interpreted with more caution. The experience from trials of  $\beta$ -interferon and other immunomodulatory therapies suggests that a major impact on MRI may be associated with more modest clinical effects.<sup>22–24</sup> Because the correlation between MRI and clinical disability in MS is modest, it may be prudent to demand a relatively high level of efficacy on MRI activity in phase II trials as a requirement for proceeding to phase III clinical outcome studies, especially if the therapy under investigation is expensive and has significant side effects.

The differences we have noted between the RR and SP subgroups are of practical importance when planning treatment trials in the future. They warrant consideration in choosing the appropriate sample size and cohort. Based on the similar patterns of activity reported in small cohorts,<sup>5–11</sup> some pilot MRI studies have combined RRMS and SPMS subgroups<sup>15–22</sup>; an obvious advantage of this approach is that patient recruitment is easier, especially now that disease modifying therapies are being increasingly used, particularly in RR patients. Our present data suggest that such a combination may be problematic if only a single pretreatment scan is obtained: in this instance a substantially larger cohort of SP patients are needed, and randomisation errors which result in uneven proportions of RR and SP patients in the treated and placebo groups could lead to spurious results. If an extra run-in scan is added, combining RR and SP cohorts may be more acceptable as the differences in sample size are smaller.

When steroids were given to treat relapses during the natural history studies from which the present clinical and MRI data were taken, an attempt was often made to perform the enhanced MRI before starting treatment, to minimise any effect on MRI activity. A three day (1g/day) course of intravenous methyl

prednisolone is the usual regime for treating relapses at our centres. Uncontrolled studies suggest that intravenous methyl prednisolone causes a temporary reduction in the number of enhancing lesions for periods ranging from one week to two months.<sup>25–26</sup> However, in one of these studies there was a total of 53 new enhancing lesions among 10 patients one month after a course of 1g intravenous methyl prednisolone/day for three days.<sup>25</sup> This suggests that the effect of this particular regime on the formation of new lesions is likely to be modest and transient.

The results also have implications for our understanding of the pathophysiology of MS. As new lesions in RRMS and SPMS usually display an initial phase of gadolinium enhancement, the concept has emerged that breakdown of the blood-brain barrier is a necessary event in the development of new pathology and, by inference, ongoing clinical deterioration.<sup>27</sup> However, we found that over a half of our SP cohort had a minimal amount of enhancing lesion activity despite the fact that they were in a phase of the disease with a poorer clinical prognosis, characterised by a steady accumulation of disability.<sup>28–29</sup> This discordance between lack of enhancement and clinical progression is already well recognised in the smaller cohort of MS patients who have a progressive and non-relapsing illness from onset (primary progressive MS).<sup>7</sup> Our findings suggest that the pathophysiology of secondary progression is not necessarily dependent on blood-brain barrier abnormality at least as seen using standard dose gadolinium enhanced MRI.

To further optimise the design of MRI outcome treatment trials, more studies are needed to elucidate factors which may influence or predict MRI activity—for example, the frequency of relapses before and during the study, entry expanded disability status scale, age, sex, disease duration, and pre-existing MRI activity. Such studies will need to examine larger cohorts of patients. Some work in this area has already been published with RR patients.<sup>21</sup> We are currently performing such an analysis in SP patients.

We acknowledge the generous support of the MS Society of Great Britain and Northern Ireland and the Dutch MS Society. NT is supported by a grant from Athena Neurosciences.

- 1 Noseworthy JH, Van der Voort MK, Wong CJ, et al. Interrater variability with the expanded disability status scale (EDSS) and functional systems (FS) in a multiple sclerosis clinical trial. *Neurology* 1990;40:971–5.
- 2 Rudick RA, Antel J, Confavreux C, et al. Clinical outcomes assessment in multiple sclerosis. *Ann Neurol* 1996;40:469–79.
- 3 Whitaker JN, McFarland HF, Rudge P, et al. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Multiple Sclerosis* 1995;1:37–47.
- 4 Miller DH, Albert PS, Barkhof F, et al. Guidelines for the use of magnetic resonance in monitoring the treatment of multiple sclerosis. *Ann Neurol* 1996;39:6–16.
- 5 Isaac C, Li DKB, Genton M, et al. Multiple sclerosis: a serial study using MRI in relapsing patients. *Neurology* 1988;38:1511–5.
- 6 Willoughby EW, Grochowski E, Li DKB, et al. Serial magnetic resonance scanning in multiple sclerosis: a second prospective study in relapsing patients. *Ann Neurol* 1989;25:43–9.
- 7 Thompson AJ, Kermode AG, Wicks D, et al. Major differences in the dynamics of primary and secondary progressive multiple sclerosis. *Ann Neurol* 1991;29:53–62.
- 8 Thompson AJ, Miller DH, Youl B, et al. Serial gadolinium-enhanced MRI in relapsing-remitting multiple sclerosis of varying disease duration. *Neurology* 1992;42:60–3.

- 9 Smith ME, Stone LA, Albert PS, *et al.* Clinical worsening in multiple sclerosis is associated with increased frequency and area of gadopentate dimeglumine enhancing magnetic resonance imaging lesions. *Ann Neurol* 1993;**33**:480–9.
- 10 Thorpe JW, Kidd D, Moseley IF, *et al.* Serial gadolinium-enhanced MRI of the brain and spinal cord in early relapsing-remitting multiple sclerosis. *Neurology* 1996;**46**:373–8.
- 11 Kidd D, Thorpe JW, Kendall BE, *et al.* MR dynamics of brain and spinal cord in progressive multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1996;**60**:15–9.
- 12 McFarland HF, Frank JA, Albert PS, *et al.* Using gadolinium-enhanced magnetic resonance imaging lesions to monitor disease activity in multiple sclerosis. *Ann Neurol* 1992;**32**:758–66.
- 13 Nauta JJP, Thompson AJ, Barkhof F, Miller DH. Magnetic resonance imaging in monitoring the treatment of multiple sclerosis patients: statistical power of parallel groups and cross-over designs. *J Neurol Sci* 1994;**122**:6–14.
- 14 Truyen L, Barkhof F, Tas M, *et al.* Specific power calculations for magnetic resonance imaging (MRI) in monitoring active relapsing-remitting multiple sclerosis (MS): implications for phase II treatment trials. *Multiple Sclerosis* 1997 (in press).
- 15 van Oosten BW, Lai M, Hodgkinson S, *et al.* Treatment of multiple sclerosis with the monoclonal anti-CD4 antibody cM-T412: results of a randomised, double-blind, placebo-controlled MR-monitored phase II trial. *Neurology* 1997 (in press).
- 16 Poser CM, Paty DW, Scheinberg L, *et al.* New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983;**13**:227–31.
- 17 Lublin F, Reingold SC. Defining the course of multiple sclerosis: results of an international survey. *Neurology* 1996;**46**:907–11.
- 18 Efron B, Tibshirani R. Statistical analysis in the computer age. *Science* 1991;**253**:390–9.
- 19 Grossman RI, Gonzales-Scarano F, Atlas SW, *et al.* Multiple sclerosis: gadolinium enhancement in MR imaging. *Radiology* 1986;**169**:117–22.
- 20 Losseff NA, Webb SL, O’Riordan JI, *et al.* Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive method with potential to monitor disease progression. *Brain* 1996;**119**:701–8.
- 21 Stone LA, Smith ME, Albert PS, *et al.* Blood-brain barrier disruption on contrast-enhanced MRI in patients with mild relapsing-remitting multiple sclerosis: relationship to course, gender and age. *Neurology* 1995;**45**:1122–6.
- 22 Edan G, Miller D, Clanet M, *et al.* Therapeutic effect of mitoxantrone combined with methylprednisolone in multiple sclerosis: a randomised multi-centre study of active disease using MRI and clinical criteria. *J Neurol Neurosurg Psychiatry* 1997;**62**:112–8.
- 23 IFNB Multiple Sclerosis Study Group. Interferon  $\beta$ -1b in the treatment of multiple sclerosis: final outcome of the randomised controlled trial. *Neurology* 1995;**45**:1277–85.
- 24 Kaurssis DM, Meiner Z, Lehmann D, *et al.* Treatment of secondary progressive multiple sclerosis with the immunomodulator linomide: a double-blind, placebo-controlled pilot study with monthly magnetic resonance imaging evaluation. *Neurology* 1996;**47**:341–6.
- 25 Miller DH, Thompson AJ, Morrisey SP, *et al.* High dose steroids in acute relapses of multiple sclerosis: MRI evidence for a possible mechanism of therapeutic effect. *J Neurol Neurosurg Psychiatry* 1992;**55**:450–3.
- 26 Barkhof F, Tas MW, Frequin STFM, *et al.* Limited duration of the effect of methylprednisolone on changes on MRI in multiple sclerosis. *Neuroradiology* 1994;**36**:382–7.
- 27 Kermode AG, Thompson AJ, Tofts PS, *et al.* Breakdown of the blood-brain barrier precedes symptoms and other MRI signs of new lesions in multiple sclerosis. *Brain* 1990;**113**:1477–89.
- 28 Confavreux C, Aimard G, Devic M. Course and prognosis of multiple sclerosis assessed by computerized data processing of 349 patients. *Brain* 1980;**103**:281–300.
- 29 Runmarker B, Andersen O. Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. *Brain* 1993;**116**:117–34.