

Clinical trials of multiple sclerosis monitored with enhanced MRI: new sample size calculations based on large data sets

M P Sormani, D H Miller, G Comi, F Barkhof, M Rovaris, P Bruzzi, M Filippi

Abstract

Objective—A new parametric simulation procedure based on the negative binomial (NB) model was used to evaluate the sample sizes needed to achieve optimal statistical powers for parallel groups (with (PGB) and without (PG) a baseline correction scan). It was also used for baseline versus treatment (BVT) design clinical trials in relapsing-remitting (RR) and secondary progressive (SP) multiple sclerosis (MS), when using the number of new enhancing lesions seen on monthly MRI of the brain as the measure of outcome.

Methods—MRI data obtained from 120 untreated patients with RRMS selected for the presence of MRI activity at baseline, 66 untreated and unselected patients with RRMS, and 81 untreated and unselected patients with SPMS were fitted using an NB distribution. All these patients were scanned monthly for at least 6 months and were all from the placebo arms of three large scale clinical trials and one natural history study. The statistical powers were calculated for durations of follow up of 3 and 6 months.

Results—The frequency of new enhancing lesions in patients with SPMS was lower, but not significantly different, from that seen in unselected patients with RRMS. As expected, enhancement was more frequent in patients with RRMS selected for MRI activity at baseline than in the other two patient groups. As a consequence, the estimated sample sizes needed to detect treatment efficacy in selected patients with RRMS were smaller than those of unselected patients with RRMS and those with SPMS. Baseline correction was also seen to reduce the sample sizes of PG design trials. An increased number of scans reduced the sample sizes needed to perform BVT trials, whereas the gain in power was less evident in PG and PGB trials.

Conclusion—This study provides reliable estimates of the sample sizes needed to perform MRI monitored clinical trials in the major MS clinical phenotypes, which should be useful for planning future studies.

(*J Neurol Neurosurg Psychiatry* 2001;70:494–499)

Keywords: multiple sclerosis; magnetic resonance imaging; sample size calculations; treatment trials

At present, the number of new gadolinium enhancing lesions on monthly MRI of the brain is one of the most used measures of outcome to monitor multiple sclerosis (MS) activity.^{1–4} New MS enhancing lesions are, however, not normally distributed and, therefore, standard approaches for sample size calculations are not desirable. Considerable effort has already been devoted to deal with the issue of sample size calculations for MRI monitored clinical trials in MS. The first paper on this topic was by Nauta *et al.*,⁵ who proposed an algorithm based on a non-parametric resampling procedure. This statistical methodology, although limited by a significant overestimation of power,⁶ was adopted in three subsequent studies.^{7–9} More recently, a parametric model based on the negative binomial (NB) distribution has been proposed⁶ to describe better than the conventional model (the Poisson model) the distribution of new enhancing lesions across patients with MS. The NB distribution is useful for modelling counts in all those situations in which the Poisson model is not able to account for a large variability of the data. In addition to offering relatively easy computation and interpretability of the data,⁶ the NB distribution model allows a better fitting of the raw data,⁶ thus giving the possibility of using parametric tests to assess new treatment efficacy and to have a more powerful tool for the sample size simulations.

Another limitation of previous work is the relatively few patients studied,^{5–8} which inevitably results in less reliable estimates.⁹ To increase the sizes of their samples, some of the previous investigators considered together patients with relapsing-remitting MS (RRMS) and secondary progressive MS (SPMS).⁵ This is again a relevant limitation, as virtually all the modern clinical trials treat individual MS phenotypes separately.^{10,11} This is based on the increasing perception that different factors are responsible for determining the clinical manifestations of the disease in the different clinical phenotypes.^{12,13} As a consequence, it is likely that in the near future different treatment options will be tested in different MS phenotypes, thus suggesting the need for calculating sample sizes for RRMS and SPMS separately.

Earlier studies^{5–9} had a limited availability of patients because they were based on data coming from natural history studies conducted in one or only a few centres. This is not the present situation any more, as several large scale, placebo controlled, MRI monitored clinical trials have been conducted on patients

Neuroimaging
Research Unit,
Department of
Neuroscience,
Scientific Institute
Ospedale San Raffaele,
University of Milan,
Via Olgettina 60, 20132
Milan, Italy

M P Sormani
M Rovaris
M Filippi

Clinical Trials Unit
G Comi

Unit of Clinical
Epidemiology and
Trials, National
Institute for Cancer
Research, Genoa, Italy
M P Sormani
P Bruzzi

NMR Research Unit,
Institute of Neurology,
London, UK
D H Miller

Dutch MS/MR Center,
Free University
Hospital, Amsterdam,
The Netherlands
F Barkhof

Correspondence to:
Dr M Filippi
filippi.massimo@hsr.it

Received 15 May 2000 and
in revised form
18 August 2000
Accepted 29 November 2000

Table 1 Numbers of new enhancing lesions/patient seen in the three patient groups studied

Number of scans	Group A			Group B			Group C		
	No of patients	Mean (SD)	Min-max	No of patients	Mean (SD)	Min-max	No of patients	Mean (SD)	Min-max
3	66	5.9 (11.6)	0-72	120	8.7 (13.9)	0-84	78	5.6 (8.4)	0-34
6	66	13.0 (20.6)	0-122	115	16.8 (22.4)	0-109	78	10.2 (14.6)	0-66

Group A: patients with RRMS not selected for MRI activity at entry;

Group B: patients with RRMS selected for MRI activity at entry;

Group C: patients with SPMS not selected for MRI activity at entry.

For further details and statistical analysis: see text.

with RRMS^{10 14-18} and SPMS.^{11 19-21} In this study, we used data of the patients with RRMS and those with SPMS enrolled in the placebo arms of three of these trials^{18 19 21} as well as those from a relatively large group of patients participating in a European multicentre natural history study²² to provide sample size calculations for RRMS and SPMS separately, which, as they are based on large datasets and on a parametric procedure, should provide more reliable guidance for planning future clinical trials in MS.

Material and methods

PATIENTS

Our datasets consisted of the following groups of patients with MS with monthly MRI of the brain for a period of at least 6 months.

Group A

This group comprised 66 patients with RRMS not selected for the presence of MRI activity at entry. There were 49 women and 17 men, with a mean age of 35.1 (SD 9.0) years, a mean duration of the disease of 6.0 (SD 5.1) years, and a mean baseline expanded disability status scale (EDSS) score²³ of 3.2 (SD 2.6).

Group B

This group consisted of 120 patients with RRMS selected for having at least one enhancing lesion at entry. There were 87 women and 33 men, with a mean age of 34.0 (SD 7.5) years, a mean duration of disease of 4.9 (SD 3.8) years, and a mean baseline EDSS of 2.4 (SD1.2).

Group C

This group comprised 81 patients with SPMS not selected for the presence of MRI activity at entry. There were 40 women and 41 men, with a mean age of 40.1 (SD 8.1) years, a mean duration of the disease of 13.4 (SD 7.3) years, and a mean baseline EDSS of 5.3 (SD1.1).

MRI

Serial dual echo conventional or fast spin echo (TR=2000-3500; TE=20-50/60-100) and postcontrast T1 weighted (TR=400-700, TE=5-25) images of the brain were acquired every month for at least six consecutive occasions. T1 weighted scans were acquired five to 10 minutes after the injection of 0.1 mmol/kg of gadolinium-DTPA. Slices were always axial and contiguous with slice thickness of 3 or 5 mm. The pixel size was about 1×1 mm for all the scans. For each patient, the same MRI acquisition methodology was kept constant throughout the entire study duration.

Scanners were not changed or upgraded over the study duration. New enhancing lesions were counted by a group of experienced observers, who always analyzed the entire set of scans from each patient.

STATISTICAL METHODOLOGY

The outcome variable of this study was the number of new enhancing lesions seen on three or six consecutive MR scans. Treatment effect was always expressed as a percentage difference between “treated” and “untreated” patients. We included in the analysis only patients with a complete 6 month follow up or with, at maximum, one missing observation (they were 66 from group A, 115 from group B, and 78 from group C). Missing observations were managed by assuming that the new enhancing lesions possibly present on the missing scans were all seen on the next subsequent scans (the total number of new enhancing lesions was obtained by summing up lesions seen during the $n-1$ months and the observation period was n months). The mean numbers of new enhancing lesions seen over the follow up period in the three groups of patients were compared using the Kruskal-Wallis and the Wilcoxon rank sum tests. The sample sizes needed for MR monitored treatment trials in MS were calculated for each of the three patient groups for a parallel groups (PG) design, for a parallel group with a single baseline correction scan (PGB) design, and for a baseline versus treatment (BVT) design (BVT design is based on the acquisition from the same patients of two sets of scans, one before and one after the introduction of the treatment to be tested and, as a consequence, each patient serves as his or her own control), using a parametric simulation method based on the NB distribution. A brief description of the statistical properties of the NB model, the estimated NB parameters obtained for the present study, and a description of the simulation algorithm used are all reported in the appendix and in another paper.⁶

Results

In table 1, the mean numbers of new enhancing lesions in each of the three patient groups are reported for 3 and 6 monthly MRI. The three patient groups had significantly different numbers of new enhancing lesions ($p=0.01$). At post hoc analysis, the difference of the mean numbers of new enhancing lesions was not statistically different between unselected patients with RRMS or SPMS, whereas both of these groups had fewer new enhancing lesions than those with RRMS selected for MRI activity at

Table 2 Numbers of patients/arm needed to perform PG trials with statistical powers of 80% or 90% to detect treatment effects ranging from 50% to 80%

Treatment effects	No of scans	Statistical power					
		80%			90%		
		Group A	Group B	Group C	Group A	Group B	Group C
50%	3	125	85	152	150	100	182
	6	118	70	141	140	85	162
60%	3	75	55	100	98	70	116
	6	65	42	94	85	55	112
70%	3	48	34	60	64	50	80
	6	40	27	58	58	35	73
80%	3	28	20	36	40	30	49
	6	24	16	34	35	22	45

Group A: patients with RRMS not selected for MRI activity at entry;
Group B: patients with RRMS selected for MRI activity at entry;
Group C: patients with SPMS not selected for MRI activity at entry.
For further details and statistical analysis: see text.

Table 3 Numbers of patients/arm needed to perform PGB trials with statistical powers of 80% or 90% to detect treatment effects ranging from 50% to 80%

Treatment effects	No of scans	Statistical powers					
		80%			90%		
		Group A	Group B	Group C	Group A	Group B	Group C
50%	3	57	43	65	80	65	85
	6	50	38	62	70	55	80
60%	3	42	32	48	58	45	62
	6	38	25	42	50	36	57
70%	3	34	26	40	47	38	52
	6	29	22	36	38	32	46
80%	3	25	19	30	35	28	40
	6	22	18	27	30	25	38

Group A: patients with RRMS not selected for MRI activity at entry;
Group B: patients with RRMS selected for MRI activity at entry;
Group C: patients with SPMS not selected for MRI activity at entry.
For further details and statistical analysis: see the text.

Table 4 Numbers of patients/arm needed to perform BVT trials with statistical powers of 80% or 90% to detect treatment effects ranging from 20% to 50%

Treatment effects	No of scans*	Statistical powers					
		80%			90%		
		Group A	Group B	Group C	Group A	Group B	Group C
20%	3	80	60	82	108	75	100
	6	48	35	62	60	52	72
30%	3	42	30	50	50	40	67
	6	22	21	35	32	32	44
40%	3	26	19	29	35	28	40
	6	14	12	20	20	18	28
50%	3	18	12	21	25	18	29
	6	12	9	16	17	12	21

*These are numbers of scans for each of the portions of the trial.
Group A: patients with RRMS not selected for MRI activity at entry;
Group B: patients with RRMS selected for MRI activity at entry;
Group C: patients with SPMS not selected for MRI activity at entry.
For further details and statistical analysis: see text.

entry ($p=0.04$ *v* unselected patients with RRMS and $p=0.001$ *v* patients with SPMS) for both durations of follow up.

The results of the power simulations are displayed in fig 1 (for a six scan follow up only), where data are fitted with exponential curves for different magnitudes of the treatment effect to allow sample size extrapolations for different experimental conditions. The numbers of patients in each treatment arm necessary to

obtain statistical powers of 80% or 90% are presented separately for the three patient groups in tables 2–4. Eighty per cent and 90% are the standard levels of statistical powers usually required in MS clinical trials. Treatment effects ranging from 50% to 80% are reported in tables 2 and 3 for PG and PGB trials, and treatment effects ranging from 20% to 50% are reported in table 4 for BVT trials. These different ranges of treatment effects were decided a priori on the basis of what conventional wisdom considers as clinically relevant and sensible expectations for the different clinical trial designs.

As expected, for trials with patients with RRMS selected for having MRI activity at entry smaller sample sizes were required than for those with unselected RRMS and SPMS for all the tested study designs. For the PG design, the sample sizes needed for treatment trials of patients with RRMS selected for MRI activity at entry were about 70% (range 63%–78%) of those needed for similar trials of unselected patients with RRMS, and about 50% (range 46%–64%) of those needed for similar trials of unselected patients with SPMS (table 2). The difference between the clinical subgroups was smaller for PGB and BVT trials: for both these designs, the sample sizes needed for treatment trials of patients with RRMS and MRI activity at entry were on average about 75% (range 66%–84%) of those needed for similar trials of unselected patients with RRMS and about 65% (range 60%–77%) of those needed for similar trials of unselected patients with SPMS (tables 3 and 4).

The baseline correction (PGB), as previously reported,^{5 8 9} reduced the sample size of PG design trials, both for patients with RRMS and those with SPMS; however the gain in power decreased as the magnitude of the expected treatment effect increased. For instance, for a treatment effect of 50%, a PGB design required about 50% fewer patients than a similar PG trial; for a treatment effect of 60%, 40% fewer patients were required; and for a treatment effect of 70%, only 20% less patients were needed. In case of an 80% treatment effect, the baseline correction resulted in virtually no gain (tables 2 and 3).

As expected, BVT trials required far fewer patients than PG or PGB trials (tables 2, 3 and 4). The BVT trials were sensitive to duration of follow up (table 4). For instance, to achieve an 80% power with a treatment effect of 20%, 80 or 48 patients with RRMS not selected for baseline activity, 60 or 35 patients with RRMS selected for activity at entry, and 80 or 60 patients with SPMS were needed if 3 or 6 monthly periods are employed.

Discussion

At present, virtually all clinical trials in MS use measures derived from enhanced MRI as primary or secondary outcomes to monitor treatment efficacy.^{10 11 14–20} As a consequence, a great deal of work has been spent in the past few years to achieve reliable estimates of the sample sizes and durations of follow up needed to perform enhanced MRI monitored clinical

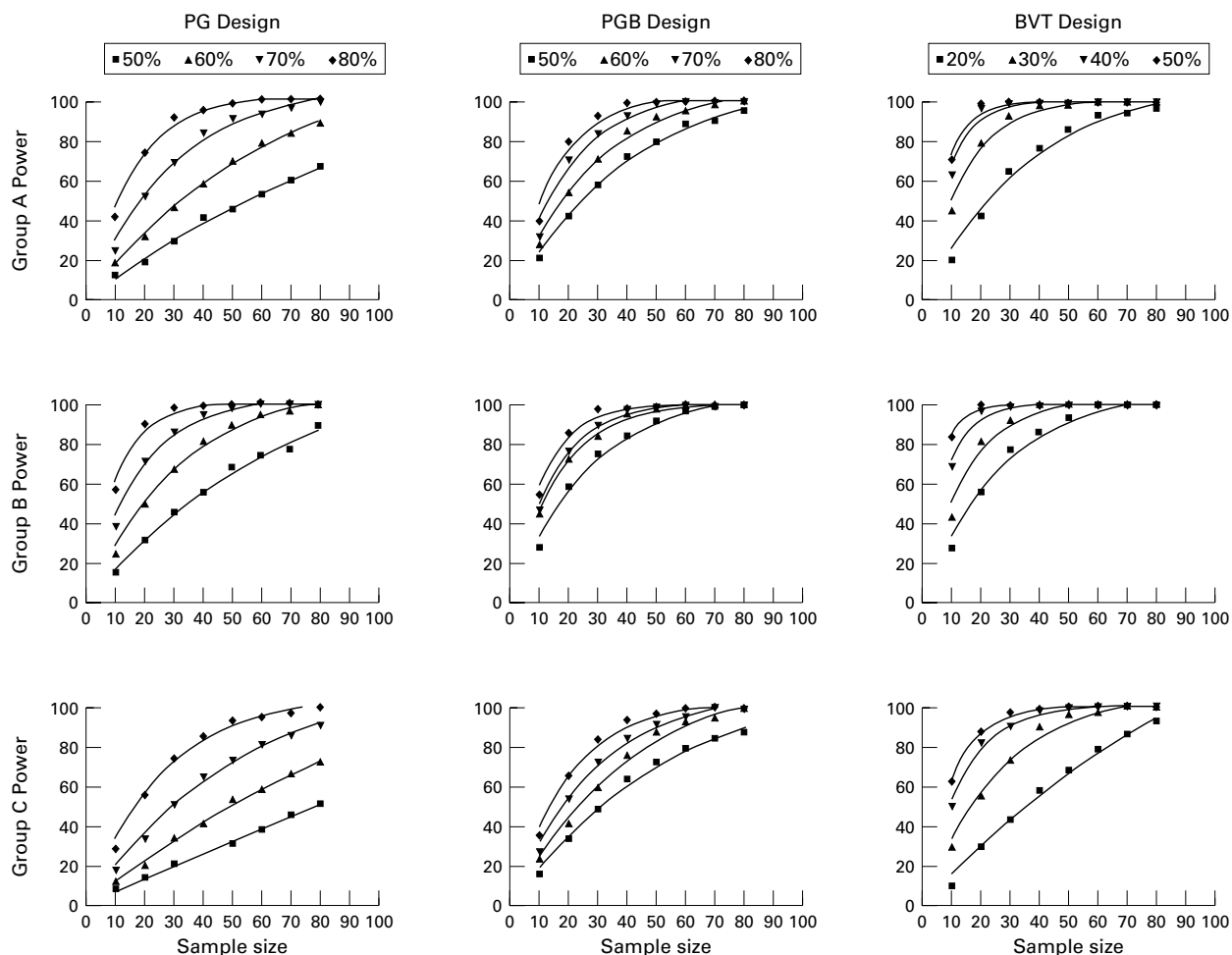


Figure 1 Power estimates for parallel group (PG) (first column), parallel group with a baseline correction scan (PGB) (second column) and baseline versus treatment (BVT) (third column) design clinical trials of unselected patients with RRMS (group A, first row), patients with RRMS and MRI activity at study entry (group B, second row) and unselected patients with SPMS (Group C, third row) followed with 6 monthly enhanced MRI. The number of new enhancing lesions is used as the measure of outcome. Raw data are fitted with exponential curves for different magnitudes of the treatment effect to allow extrapolations for different trial designs.

trials in MS with adequate statistical powers in different experimental conditions.⁵⁻⁹ Overestimations and underestimations of the numbers of patients and MRI needed to reject the null hypothesis with adequate powers are undesirable. On the one hand, patient and scan overestimations would result in unnecessarily high trial costs and patient inconvenience. On the other, patient and scan underestimations would lead to equivocal or even false negative trial results. In the past, when large serial MRI datasets were not available, power calculations were based on data from small samples of patients with MS,⁵⁻⁹ and, as a consequence, they were relatively unreliable⁹ and potentially inaccurate.⁶ More recently, the situation has changed for two reasons. Firstly, large serial MRI data sets coming from the placebo arms of several MS clinical trials are available.¹⁴⁻²¹ Secondly, a new modelling of MS lesion counts, which results in more accurate power estimates than the conventional approach, has been developed to study patients with MS.⁶

Against this novel background, we performed the present study to provide more reliable and accurate estimates of the sample sizes needed to perform MRI monitored clinical trials with adequate statistical power in patients

with different MS phenotypes and for different trial designs, when the number of new enhancing lesions is the measure of outcome. The tables and fig 1 of the present paper should provide a valuable reference when planning future clinical trials in MS, as they are based on large MRI datasets and on a reliable parametric simulation procedure.⁶ Such an approach might become even more valuable in the near future, as it is likely that new experimental treatments will be tested against already approved treatments. In this scenario, the sample sizes needed to detect treatment effects should increase dramatically and adequate sample estimations would reduce the risks of running uninformative or extremely expensive trials.

This study also confirms that including MRI activity at study entry as a patient selection criterion allows a substantial reduction of patients needed to run a PG design trial. The presence of MRI activity at a given time point is associated with a higher likelihood of further MRI activity during the next few months.²⁴ Thus, by reducing the number of uninformative (inactive) scans, smaller sample sizes and shorter durations of follow up will be required to achieve enough study power. Nevertheless,

such an approach has a hidden cost, as it is likely that a significant number of clinically eligible patients would be screened, but on showing a negative enhanced scan, would not enter the trial. The major drawback of such an approach is, however, the increased risk of selecting patients during periods of relatively high disease activity which will inevitably decline during the period of the trial regardless of the treatment used. This effect is known as “regression to the mean” and its impact is proportional to the amount of activity required for the patients’ eligibility and might be misleading if the design of the study lacks a control arm. In addition, the use of selection criteria makes the results of a trial less generalisable to all the potential patients that might benefit from the treatment.

In PG trials, the use of additional MRI for baseline correction, which reduces the between patient variability of new enhancing lesions, also results in smaller sample sizes than those needed for PG trials conducted without such an approach. This agrees with findings of previous studies^{5,7,8} and with recent work demonstrating that the addition of a second and sometimes of a third baseline scan results in a significantly reduced sample variance (A Smith, J Petkau, personal communication). Although the use of additional baseline scans may not always be ethical or feasible, as there might be concerns about withholding a treatment from patients for a certain period of time or about the additional costs, the addition of a single baseline scan 1 month earlier is unlikely to be problematic. Indeed, cost considerations are mitigated, as shown in the present study, by the fact that the acquisition of a baseline pretreatment scan substantially reduces the number of scans needed during the course of the study.

Although the algorithm used in the present study to perform power calculations was shown to provide a better fitting of the raw data than the previously used methodology,⁶ it is not without potential problems. The main assumption behind our calculations is that a given experimental treatment would be able to modify the number of new enhancing MS lesions, but not their distribution, which should always be fitted by the NB model. This assumption can only be verified by analyzing the data of the treated arms of clinical trials and further work is warranted to clarify this issue. Another caveat of any of the possible approaches to perform power calculations is that the sizes of the samples studied and the durations of the follow up periods should not only be determined by statistical considerations, but should also be based on the characteristic of the experimental treatment used. Finally, the well known limitation in the correlation of enhancing lesion activity with clinical measures, such as relapse rate and disability,²⁴ is still widely regarded as justifying the use of the latter as the primary end points in definitive phase III clinical trials.

We are very grateful to TEVA Pharmaceutical Industries Ltd, Schering AG, and Elan Pharmaceuticals Ltd for providing us

with the data of the placebo arms of previous clinical trials in MS. We are also indebted to Drs P D Molyneux and N Tubridy for their help in collecting some of the MRI data.

Appendix

THE NEGATIVE BINOMIAL (NB) MODEL

The NB distribution is of central importance within the family of the so called “mixed Poisson distributions”, because of its convenient mathematical properties. An NB model is a Poisson process in which the mean is random and follows a γ distribution. A mixed Poisson distribution has the following probabilities:

$$p_k = \int_0^{\infty} \frac{\mu^k \cdot e^{-\mu}}{k!} \cdot u(\mu) d\mu \quad k = 0, 1, 2, \dots$$

If $u(x)$ is the probability density function for the γ distribution with parameter ϑ :

$$u(x) = \left(\frac{\vartheta}{\Gamma(\vartheta)} \right) \cdot (\vartheta \cdot x)^{(\vartheta-1)} \cdot e^{-(\vartheta \cdot x)}$$

then, the mixed distribution is NB with parameters μ and ϑ :

$$p_k = \frac{\Gamma(k + \vartheta)}{\Gamma(\vartheta) \cdot k!} \cdot \frac{\vartheta^\vartheta \cdot \mu^k}{(\vartheta + \mu)^{(\vartheta+k)}}$$

It can be easily verified that the NB distribution has $E(k) = \mu$ and $Var(k) = \mu + \mu^2/\vartheta$.

NB PARAMETER ESTIMATES FOR THE PRESENT STUDY

In the present study, the μ and ϑ estimated for the distribution of new enhancing lesions seen on monthly MRI using the NB model are reported in the following table for the three clinical subgroups studied and for different durations of follow up (3 or 6 months).

The values for μ and ϑ reported in the above table are those which fit, according to the NB model, the actual data reported in table 1 of the main text. The mean numbers of new enhancing lesions/patient seen for each of the three patient groups studied is estimated by μ , whereas the corresponding SDs are estimated by taking the square root of $\mu + \mu^2/\vartheta$. The NB distributions better fitting the actual numbers of new enhancing lesions seen in the three different patient groups are shown in the figure.

THE SIMULATION PROCEDURE

The parameters μ and ϑ obtained by fitting the NB model to each set of the collected MRI data were considered representative of the untreated MS patient group used to estimate them. For each trial, the “untreated” group was obtained by randomly sampling from an NB distribution with these parameters. For the PG design, the “treated” group was obtained by randomly sampling from an NB distribution with parameters $\mu_{treated} = \mu_{untreated} \cdot (1 - \text{treatment effect})$ and $\vartheta_{treated} = \vartheta_{untreated}$ (the treatment was supposed to leave ϑ unchanged). The baseline correction scan was simulated by random sampling from a Poisson distribution with parameter generated for each patient by sampling from a γ distribution. The parameter of the γ distribution was the obtained by the NB fitting. The utility of the baseline correction lies in the fact that such

No of scans	Group A		Group B		Group C	
	μ	ϑ	μ	ϑ	μ	ϑ
3	5.9	0.49	8.7	0.62	5.6	0.39
6	13.0	0.52	16.8	0.75	10.2	0.37

Group A: patients with RRMS not selected for MRI activity at entry;

Group B: patients with RRMS selected for MRI activity at entry;

Group C: patients with SPMS not selected for MRI activity at entry.

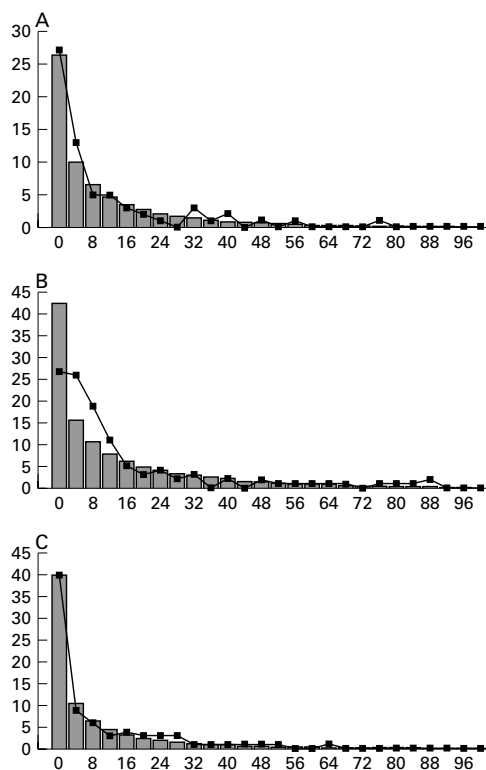


Figure 2 Numbers of observed new enhancing lesions (black squares connected by lines) and numbers of those expected under the NB model assumptions (bars) for patients with RRMS (A) or SPMS (C) not selected for the presence of MRI activity at entry, and for patients with RRMS selected for having at least one enhancing lesion at entry (B).

a correction, by reducing the between patient variability of new enhancing lesions, enables the conduction of PG trials with smaller sample sizes than those needed when it is not used. For the BVT design, each patient was simulated by sampling from a γ distribution with parameter $\vartheta_{untreated}$ and expected value equal to $\mu_{untreated}$. For each patient (i), the number of lesions counted over the baseline period was simulated by randomly sampling from a Poisson distribution with parameter $\mu_{untreated}$ and the number of lesions counted over the treatment period was simulated by randomly sampling from a Poisson distribution with parameter $\mu_{untreated} * (1 - \text{treatment effect})$. For each experimental design, 1000 trials were generated and the power was calculated as the proportion of trials which yielded a significant result. The effect of treatment was assessed using a Wilcoxon test.

- 1 Miller DH, Barkhof F, Nauta JJP. Gadolinium enhancement increased the sensitivity of MRI in detecting disease activity in MS. *Brain* 1993;116:1077–94.
- 2 McFarland HF, Frank JA, Albert PS, et al. Using gadolinium-enhanced magnetic resonance imaging to monitor disease activity in multiple sclerosis. *Ann Neurol* 1992;32:758–66.
- 3 Miller DH, Albert PS, Barkhof F, et al. Guidelines for the use of magnetic resonance techniques in monitoring the treatment of multiple sclerosis. *Ann Neurol* 1996;39:6–16.

- 4 Rovaris M, Filippi M. Magnetic resonance techniques to monitor disease evolution and treatment trial outcomes in multiple sclerosis. *Curr Opin Neurol* 1999;12:337–44.
- 5 Nauta JJP, Thompson AJ, Barkhof F, et al. Magnetic resonance imaging in monitoring the treatment of multiple sclerosis patients, statistical power of parallel-groups and crossover designs. *J Neurol Sci* 1994;122:6–14.
- 6 Sormani MP, Bruzzi P, Miller DH, et al. Modelling MRI enhancing lesion counts in multiple sclerosis using a negative binomial model: implications for clinical trials. *J Neurol Sci* 1999;163:74–80.
- 7 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 therapeutic trials. *Multiple Sclerosis* 1997;2:283–90.
- 8 Tubridy N, Ader HJ, Barkhof F, et al. Exploratory treatment trials in multiple sclerosis using MRI: sample size calculations for relapsing remitting and secondary progressive subgroups using placebo controlled parallel groups. *J Neurol Neurosurg Psychiatry* 1998;64:50–5.
- 9 Sormani MP, Molyneux PD, Gasperini C, et al. Statistical power of MRI monitored trials in multiple sclerosis: new data and comparison with previous results. *J Neurol Neurosurg Psychiatry* 1999;66:465–9.
- 10 PRISMS Study Group. Randomised double-blind placebo-controlled study of interferon β -1a in relapsing-remitting multiple sclerosis. *Lancet* 1998;352:1498–504.
- 11 European Study Group on Interferon β -1b in Secondary Progressive MS. Placebo-controlled multicentre randomised trial of interferon β -1b in treatment of secondary progressive multiple sclerosis. *Lancet* 1998;352:1491–7.
- 12 Filippi M, Iannucci G, Tortorella C, et al. Comparison of MS clinical phenotypes using conventional and magnetization transfer MRI. *Neurology* 1999;52:588–94.
- 13 Tortorella C, Viti B, Bozzali M, et al. A magnetization transfer histogram study of normal appearing brain tissue in multiple sclerosis. *Neurology* 2000;54:186–93.
- 14 Simon JH, Jacobs LD, Campion M, et al. Magnetic resonance studies of intramuscular interferon β -1a for relapsing multiple sclerosis. *Ann Neurol* 1998;43:79–87.
- 15 The Once Weekly Interferon for MS Study Group (OWIMS). Evidence of interferon β -1a dose response in relapsing-remitting MS. The OWIMS study. *Neurology* 1999;53:679–86.
- 16 Li DKB, Paty DW, the UBC MS/MRI Analysis Research Group, the PRISMS Study Group. Magnetic resonance imaging results of the PRISMS trial: A randomized, double-blind, placebo-controlled study of interferon β -1a in relapsing-remitting multiple sclerosis. *Ann Neurol* 1999;46:197–206.
- 17 Sorensen PS, Wanscher B, Jensen CV, et al. Intravenous immunoglobulin G reduces MRI activity in relapsing multiple sclerosis. *Neurology* 1998;50:1273–81.
- 18 Comi G, Filippi M, the Copaxone MRI Study Group. The effect of glatiramer acetate (Copaxone) on disease activity as measured by cerebral MRI in patients with relapsing-remitting multiple sclerosis (RRMS): a multi-center, randomized, double-blind, placebo-controlled study extended by open-label treatment [Abstract]. *Neurology* 1999;52(suppl 2):A289.
- 19 Miller DH, Molyneux PD, Barker GJ, et al. Effect of interferon β -1b on magnetic resonance imaging outcomes in secondary progressive multiple sclerosis: results of a European multicenter, randomized, double-blind, placebo-controlled trial. *Ann Neurol* 1999;46:850–9.
- 20 Paolillo A, Coles AJ, Molyneux PD, et al. Quantitative MRI in patients with secondary progressive MS treated with monoclonal antibody Campath 1H. *Neurology* 1999;53:751–7.
- 21 Tubridy N, Behan PO, Capildeo R, et al. The effect of anti- α 4 integrin antibody on brain lesion activity in MS. The UK Antegren Study Group. *Neurology* 1999;53:466–72.
- 22 Molyneux PD, Filippi M, Barkhof F, et al. Correlations between monthly enhanced MRI lesion rate and changes in T2 lesion volume in multiple sclerosis. *Ann Neurol* 1998;43:332–9.
- 23 Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33:1444–52.
- 24 Kappos L, Moeri D, Radue EW, et al. Predictive value of gadolinium-enhanced MRI for relapse rate and changes in disability/impairment in multiple sclerosis: a meta-analysis. *Lancet* 1999;353:964–9.