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.

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,¹ who proposed an algorithm based on a non-parametric resampling procedure. This statistical methodology, although limited by a significant overestimation of power,⁵ was adopted in three subsequent studies.⁶⁻⁸ More recently, a parametric model based on the negative binomial (NB) distribution has been proposed⁹ 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,⁴ the NB distribution model allows a better fitting of the raw data,⁹ 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,⁶⁻⁸ which inevitably results in less reliable estimates.⁷ 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.¹²⁻¹¹ 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 studies⁶⁻⁹ 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...
Table 1  Numbers of new enhancing lesions/patient seen in the three patient groups studied

<table>
<thead>
<tr>
<th>Number of scans</th>
<th>Group A</th>
<th></th>
<th>Group B</th>
<th></th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of patients</td>
<td>Mean (SD)</td>
<td>Min-max</td>
<td>No of patients</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
<td>5.9 (11.6)</td>
<td>0–72</td>
<td>120</td>
<td>8.7 (13.9)</td>
</tr>
<tr>
<td>6</td>
<td>66</td>
<td>13.0 (20.6)</td>
<td>0–122</td>
<td>115</td>
<td>16.8 (22.4)</td>
</tr>
</tbody>
</table>

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 and SPMS. 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 as well as those from a relatively large group of patients participating in a European multicentre natural history study 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) score 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 (SD 1.2).

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 13.8 (SD 7.3) years, and a mean baseline EDSS of 5.3 (SD 1.1).

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 1x1 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.

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 entry.
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%.

<table>
<thead>
<tr>
<th>Treatment effects</th>
<th>Statistical power</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of scans</td>
<td>Group A</td>
<td>Group B</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>3</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>118</td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>65</td>
</tr>
<tr>
<td>70%</td>
<td></td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>40</td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>24</td>
</tr>
</tbody>
</table>

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%.

<table>
<thead>
<tr>
<th>Treatment effects</th>
<th>Statistical power</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of scans</td>
<td>Group A</td>
<td>Group B</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>3</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>38</td>
</tr>
<tr>
<td>70%</td>
<td></td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>22</td>
</tr>
</tbody>
</table>

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 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%.

<table>
<thead>
<tr>
<th>Treatment effects</th>
<th>Statistical power</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of scans*</td>
<td>Group A</td>
<td>Group B</td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td>3</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>48</td>
</tr>
<tr>
<td>30%</td>
<td></td>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>

*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.

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**: 50%, 60%, 70%, 80%
- **Statistical powers**: 80%, 90%
- **Groups**: 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)

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**: 50%, 60%, 70%, 80%
- **Statistical powers**: 80%, 90%
- **Groups**: 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)

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**: 20%, 30%, 40%, 50%
- **Statistical powers**: 80%, 90%
- **Groups**: 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)

Discussion:

At present, virtually all clinical trials in MS use measures derived from enhanced MRI as primary or secondary outcomes to monitor treatment efficacy. 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 trials.
trials in MS with adequate statistical powers in different experimental conditions.\textsuperscript{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,\textsuperscript{5–9} and, as a consequence, they were relatively unreliable\textsuperscript{9} and potentially inaccurate.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 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, 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, is still widely regarded as justifying the use of the latter as the primary end points in definitive phase III clinical trials.

The data of the placebo arms of previous clinical trials in MS. We are also indebted to Des P-D Moloney 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 Gamma distribution. A mixed Poisson distribution has the following probabilities:

$$P_k = \frac{\mu^k \cdot \theta^k}{k!} \cdot u(\mu) d\mu$$

If $u(x)$ is the probability density function for the Gamma distribution with parameter $\theta$:

$$u(x) = \frac{\theta^x \cdot (\theta - x)^{\omega - 1} \cdot e^{\omega x}}{\Gamma(\omega)}$$

then, the mixed distribution is NB with parameters $\mu$ and $\theta$:

$$P_k = \frac{\Gamma(k + \theta) \cdot \theta^k \cdot \mu^k}{\Gamma(\omega) \cdot k!} \cdot u(\mu) d\mu$$

It can be easily verified that the NB distribution has

$$E(k) = \mu + \omega \mu^2 / \theta$$

NB parameter estimates for the present study

In the present study, the $\mu$ and $\theta$ 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 $\mu$ and $\theta$ 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 $\mu$, whereas the corresponding SDs are estimated by taking the square root of $\omega \mu^2 / \theta$. 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 $\mu$ and $\theta$ 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 $\mu_{treated} = \mu_{untreated} \cdot (1 - treatment \ effect)$ and $\theta_{treated} = \theta_{untreated}$ (the treatment was supposed to leave $\theta$ unchanged). The baseline correction scan was simulated by random sampling from a Poisson distribution with parameter generated for each patient by sampling from a Gamma distribution. The parameter of the $\Gamma$ distribution was the obtained by the NB fitting. The utility of the baseline correction lies in the fact that such

<table>
<thead>
<tr>
<th>No of scans</th>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$\mu$</td>
</tr>
<tr>
<td>3</td>
<td>5.9</td>
<td>0.49</td>
<td>8.7</td>
</tr>
<tr>
<td>6</td>
<td>13.0</td>
<td>0.52</td>
<td>16.8</td>
</tr>
</tbody>
</table>

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.

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.
Clinical trials of multiple sclerosis monitored with enhanced MRI


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).
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 and M Filippi

J Neurol Neurosurg Psychiatry 2001 70: 494-499
doi: 10.1136/jnnp.70.4.494

Updated information and services can be found at:
http://jnnp.bmj.com/content/70/4/494

These include:

References
This article cites 24 articles, 10 of which you can access for free at:
http://jnnp.bmj.com/content/70/4/494#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Immunology (including allergy) (1943)
Multiple sclerosis (934)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/