Scoring treatment response in patients with relapsing multiple sclerosis
MP Sormani, J Rio, M Tintorè, A Signori, D Li, P Cornelisse, B Stubinski, ML Stromillo, X Montalban and N De Stefano
Mult Scler 2013 19: 605 originally published online 25 September 2012
DOI: 10.1177/1352458512460605
The online version of this article can be found at:
http://msj.sagepub.com/content/19/5/605

Published by:
http://www.sagepublications.com

Additional services and information for Multiple Sclerosis Journal can be found at:

Email Alerts: http://msj.sagepub.com/cgi/alerts
Subscriptions: http://msj.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Apr 18, 2013
OnlineFirst Version of Record - Sep 25, 2012

What is This?
Scoring treatment response in patients with relapsing multiple sclerosis

MP Sormani¹, J Rio², M Tintorè², A Signori¹, D Li³, P Cornelisse⁴, B Stubinski⁴, ML Stromillo⁵, X Montalban²* and N De Stefano⁵*

Abstract

Background: We employed clinical and magnetic resonance imaging (MRI) measures in combination, to assess patient responses to interferon in multiple sclerosis.

Objective: To optimize and validate a scoring system able to discriminate responses to interferon treatment in patients with relapsing–remitting multiple sclerosis (RRMS).

Methods: Our analysis included two large, independent datasets of RRMS patients who were treated with interferons that included 4-year follow-up data. The first dataset (“training set”) comprised of 373 RRMS patients from a randomized clinical trial of subcutaneous interferon beta-1a. The second (“validation set”) included an observational cohort of 222 RRMS patients treated with different interferons. The new scoring system, a modified version of that previously proposed by Rio et al., was first tested on the training set, then validated using the validation set. The association between disability progression and risk group, as defined by the score, was evaluated by Kaplan Meier survival curves and Cox regression, and quantified by hazard ratios (HRs).

Results: The score (0–3) was based on the number of new T2 lesions (>5) and clinical relapses (0, 1 or 2) during the first year of therapy. The risk of disability progression increased with higher scores. In the validation set, patients with a score of 0 showed a 3-year progression probability of 24%, while those with a score of 1 increased to 33% (HR = 1.56; p = 0.13), and those with score greater than or equal to 2 increased to 65% (HR = 4.60; p < 0.001).

Conclusions: We report development of a simple, quantitative and complementary tool for predicting responses in interferon-treated patients that could help clinicians make treatment decisions.

Keywords
Interferon, patient assessment, relapsing–remitting multiple sclerosis, scoring system, treatment response

Date received: 24th June 2012; accepted: 12th August 2012

Introduction

The response of multiple sclerosis (MS) patients to disease-modifying drugs (DMDs) is very heterogeneous, which makes it difficult to establish whether, and to what degree, one treatment is not producing the desired effect.¹ However, because the range of available treatments is broadening, it is becoming imperative for clinicians to accurately predict an individual patient’s response to a given therapy, in order to make optimal therapeutic decisions.

The assessment of short-term responses to therapy relies on outcome measures that may directly or indirectly reflect disease progression. Both clinical and magnetic resonance imaging (MRI) measures have proven useful in detecting disease activity and progression in patients with the relapsing–remitting form of MS (RRMS) who are treated with DMDs, and these two measures have been used in combination, to assess treatment response.²⁻⁵ In particular, Rio and colleagues⁴ proposed a scoring system based on the combined assessment at 1-year of clinical relapses, disability progression as measured by the Expanded Disability...
A recent study shows that the combined measurement of 1-year brain MRI activity with clinical relapses is a good surrogate for assessing the short-term risk of disability progression in RRMS patients who are treated with interferon (IFN). This observation provides a scientific rationale to test a scoring method for predicting responses to DMDs, based on clinical relapses and/or brain MRI metrics. We used two independent, longitudinal datasets: one from a large, randomized clinical trial of subcutaneous (sc) IFN beta-1a therapy for RRMS and one obtained from an observational cohort of RRMS subjects who were treated with IFN, with the aims to:

1. optimize a new scoring system and test its ability to identify RRMS patients who are responding/not-responding to treatment (using one dataset, the ‘training set’);
2. validate the new scoring system using another independent dataset (‘validation set’).

## Defining risk groups

To identify those patients at risk of having a poor response to treatment, we developed a simplified version of the score reported by Rio et al. From now on we will refer to this modified score as the “modified Rio score”.

**Rio score.** This scoring system is based on a combination of MRI, relapse and EDSS criteria, as defined after the first year of therapy:

(a) MRI criterion = 1; if the patient had (on the yearly MRI scan) > 2 active T2 lesions, defined as new or enlarging T2-weighted lesions, plus the number of gadolinium-enhancing (Gd) T1-weighted lesions over the first year;
(b) relapse criterion = 1; if the patient experienced ≥ 1 relapse over the first year;
(c) EDSS criterion = 1; if there was an increase in the patient’s EDSS score of ≥1 point, sustained over at least 6 months and confirmed at the end of the follow-up period.

The sum of these three criteria classifies patients into those having a score of 0, 1, 2 or 3 (from a low risk to a high risk of having a poor treatment response).

**Modified Rio score.** This is a simplified version of the Rio score, which was first developed by working on the training set. In the PRISMS dataset, new lesions occurring on proton density/T2-weighted MRI scans were counted over the first year of treatment on scans performed every 6 months. Over
the same period, the number of relapses and the EDSS changes that were confirmed at 6-month visits were assessed. Statistical modeling was used to optimize the score on the training set (see supplementary appendix). Briefly, we first modeled the relationship between the numbers of new T2 MRI lesions and the numbers of relapses with the probability of EDSS progression over the subsequent 3 years, using univariate Cox models. Because it is unlikely that the risk of disability progression increases linearly with the number of new T2 MRI lesions and the number of relapses, these two variables were transformed, using five additional different smoothing functions, and the transforms that gave the best fits were chosen (see Supplementary Appendix for details on the smoothing functions). For those transforms based on a cut-off value, the optimal cut-off was determined as the one yielding the best fit in the Cox model. After this process, the optimized score comprised the following two criteria:

(a) MRI criterion = 1; if the patient has had > 5 new T2 lesions;
(b) relapse criterion = 1; if the patient experienced 1 relapse; and relapse criterion = 2 if the patient experienced ≥ 2 relapses.

The score was created by simply summing up these two criteria, as follows:

score = 0; if new T2 lesions ≤ 5 and relapses = 0;
score = 1; if new T2 lesions ≤ 5 and relapses = 1; or new T2 lesions > 5 and relapses = 0;
score = 2; if new T2 lesions ≤ 5 and relapses ≥ 2; or new T2 lesions > 5 and relapses = 1;
score = 3; if new T2 lesions > 5 and relapses ≥ 2.

Because the number of new T2 MRI lesions was assessed on 6-month scans in the training set6-7 and on 12-month scans in the validation set, the cut-off value of 5 found in the training set was transformed to a 4 in the validation set (see Table 2). This was decided after an additional analysis verified that the number of new T2 MRI lesions that were counted on the same subjects on yearly scans is about 75% of the number of new T2 MRI lesions, if counted on 6-month scans (see Supplementary Appendix for details). Therefore, for the validation set, the scores were as follows:

score = 0; new T2 lesions ≤ 4 and relapses=0;
score = 1; new T2 lesions ≤ 4 and relapses=1, or new T2 lesions>4 and relapses=0;
score = 2; new T2 lesions ≤ 4 and relapses≥2, or new T2 lesions>4 and relapses=1;
score = 3; new T2 lesions > 4 and relapses≥2.

This procedure was then applied to test for the ability of the modified Rio score to discriminate patients with different risk levels of disability progression and of relapses (the treatment responders versus non-responders) on the training set, plus to validate the modified Rio score using the validation set.

In these analyses, according to the score, patients were classified as being at low risk (score = 0), intermediate risk (score = 1) or high risk (score = 2–3) of a poor response to treatment, after 1 year of therapy.

Outcomes

Patient prognosis under IFN treatment was assessed using 1 year from treatment initiation as the reference point. Thus, disease progression was quantified in each of the subsequent 3 years of follow-up, as follows:

Relapse rate. In both datasets, the number of clinical relapses experienced over 3 years was counted and the

<table>
<thead>
<tr>
<th>Table 2. Distribution of patients according the modified Rio score criteria.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modified Rio score on the training set (PRISMS database)</strong></td>
</tr>
<tr>
<td><strong>Score components</strong></td>
</tr>
<tr>
<td>MRI criterion</td>
</tr>
<tr>
<td>≤ 5 new T2 lesions</td>
</tr>
<tr>
<td>&gt; 5 new T2 lesions</td>
</tr>
<tr>
<td>Relapse criterion</td>
</tr>
<tr>
<td>= 0 relapses</td>
</tr>
<tr>
<td>≥ 1 relapse</td>
</tr>
<tr>
<td>≥ 2 relapses</td>
</tr>
<tr>
<td><strong>Modified Rio score</strong></td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

MRI: magnetic resonance imaging.
The average number of relapses per year estimated. The definition of clinical relapse was substantially similar in both datasets. The relapse rate during the 3 years of follow-up was compared across the risk groups, using a negative binomial regression analysis.

Disability progression. In the PRISMS dataset, disability progression was defined as an increase of $\geq 1$ EDSS point (0.5; if the 1-year EDSS was $\geq 6$), that was confirmed at 3 months; while in the Barcelona dataset, it was defined as an increase of $\geq 1$ EDSS point (0.5; if the 1-year EDSS was $\geq 6$) that was confirmed at 6 months, or at the stopping of or change in patient therapy due to a perceived treatment failure.

Patients with disability progression were defined as non-responders to IFN. Time to disability progression was calculated as the time between the end of 1 year of therapy and the visit reporting the disability progression, as defined above. The association between assessed disability progression over 3 years of follow-up and the risk groups, as defined by the modified Rio score, was evaluated by Kaplan Meier survival curves with the log-rank test (test for heterogeneity) and by Cox regression (test for trend), and then quantified by hazard ratios (HRs).

Because the probability to progress in the validation set reflected the prevalence of progression in the IFN-treated population, it was possible to estimate the probability for a patient to progress, if classified as a non-responder (positive predictive value, PPV) and the probability of a patient being free from progression, if classified as a responder (negative predictive value, NPV). Sensitivity, specificity and accuracy were also estimated.

Standard protocols of the original studies were approved by institutional review boards in the participating centers. All patients gave written, informed consent.

Results

Testing of the modified Rio scoring system

Table 2 reports the distribution of the modified Rio score in the PRISMS dataset (training set). Overall, after 1 year of IFN treatment, 40.5% of patients had a score $= 0$; while 30.7% had a score $= 1$; 23.6% had a score $= 2$; and 5.1% had a score $= 3$.

The modified Rio score assessed after 1 year of interferon therapy was deemed highly predictive of the patient’s relapse frequency in the subsequent 3 years: the average (SD: standard deviation) number of relapses over 3 years was found to be 0.99 (1.19) for those patients with a score $= 0$; while it was found to be 2.07 (1.69) for patients with a score $= 1$; and 3.14 (2.42) for patients with a score $\geq 2$ (the test for heterogeneity gave $p < 0.001$ and the test for trend gave $p < 0.001$; as can be seen in Figure 1(a)).

We found there was a close association between the probability of disease progression in the patient and increasing scores (Figure 2(a); log-rank test for heterogeneity $p = 0.009$; test for trend, HR = 1.36, 95% CI = 1.12–1.66, $p = 0.002$). Patients with a score $= 0$ had a probability to progress, in 3 years, of 32%; while those with score $= 1$ had a probability to progress, in 3 years, of 42% (HR = 1.35; 95% CI = 0.89–2.04; $p = 0.15$); those attaining a score $= 2$ had a probability to progress, in 3 years, of 46% (HR = 1.36; 95% CI = 1.12–1.66, $p = 0.002$).
1.65; 95% CI = 1.08–2.54; \( p = 0.02 \)); and lastly, those with a score = 3 had a probability to progress, in 3 years, of 71% (HR = 3.06; 95% CI = 1.62–5.78; \( p = 0.001 \)). Given the small number of patients within the highest risk category, those with scores of 2 and 3 were grouped together, to define the high risk group as patients with score \( \geq 2 \); and these were found to have a probability to progress of 50% (HR = 1.86; 95% CI = 1.25–2.76; \( p = 0.002 \)) (Figure 2(a)).

When the modified Rio score was applied to patients from the point of treatment initiation (i.e., 4-year follow-up), those with score = 0 had a probability to progress of 28%; those with a score = 1 of 46% (HR = 1.82; 95% CI = 1.21–2.73; \( p = 0.004 \)); and those with a score of \( \geq 2 \) had a probability to progress of 60% (HR = 2.95; 95% CI = 2.36–3.79; \( p < 0.001 \)) (Figure 2(b)).

There was a close association between the probability of disease progression and an increasing score (Figure 3(a); log-rank test for heterogeneity \( p < 0.001 \); test for trend, HR = 1.93, 95% CI = 1.41–2.65, \( p < 0.001 \)). Patients with a score = 0 had a probability to progress of 24%, those with a score = 1 had a probability to progress of 33% (HR = 1.56; 95% CI = 0.87–2.78; \( p = 0.13 \)), those with a score = 2 had a probability to progress of 63% (HR = 4.56; 95% CI = 2.36–8.79; \( p < 0.001 \)) and lastly, those with a score = 3 had a probability to progress of 75% (HR = 4.77; 95% CI = 1.47–15.5; \( p = 0.009 \)). Given the small number of patients in the highest risk category, the scores 2 and 3 were also grouped together, to define the highest risk group: patients with a score \( \geq 2 \) had a probability to progress of 65% (HR = 4.60; 95% CI = 2.51–8.43; \( p < 0.001 \)) (Figure 3(a)).

Also using this dataset, we found that if the score was applied to patients from the point of treatment initiation, those with a score = 0 had a probability to progress in 4 years of 20%; those with score = 1 had a probability to progress in 4 years of 43% (HR = 2.57; 95% CI = 2.47–4.50; \( p = 0.001 \)), and those with a score \( \geq 2 \) had a probability to progress in 4 years of 78% (HR = 7.02; 95% CI = 3.83–12.87; \( p < 0.001 \)), confirming that there was better discrimination among groups with longer follow-up times, as was found in the training set (Figure 3(b)).

Selecting patients with complete 4-year follow-ups from the initiation of treatment, we calculated the sensitivity, specificity, accuracy, PPV and NPV of the modified Rio scoring system in identifying patients with an EDSS progression or with > 1 relapse, during follow-up. Defining patients with a score = 2–3 as non-responders, and patients with a score = 0–1 as responders, the PPV obtained was 83% and the NPV was 68%. We found that the score has

**Validation of the modified Rio score**

The prognostic value of the modified Rio score was then tested in the independent Barcelona (“validation”) dataset. Overall, after 1 year of IFN treatment, we found that 66.2% of patients had a score = 0; while 23.4% had a score = 1; plus 8.6% had a score = 2; and 1.8% had a score = 3 (see Table 2). We found that the modified Rio score, after 1 year of therapy, was able to be highly predictive of relapse frequency within the subsequent 3 years: the average (SD) number of relapses/year, over 3 years, was 0.21 (0.27) for patients with a score = 0; as well as 0.26 (0.36) for patients with a score = 1; and 0.75 (0.56) for patients with a score \( \geq 2 \) (test for heterogeneity \( p < 0.001 \); test for trend \( p < 0.001 \); see Figure 1(b)).

![Figure 2. Probability of multiple sclerosis disability progression from the first year since interferon treatment started (3 years, (a)) and over the follow-up period (4 years, (b)), according to the modified Rio score assessment of the data in the training set (PRISMS study data).](image-url)
low sensitivity (24%) and very high specificity (97%), yielding a global accuracy of 69%.

Discussion

The early identification of patients with a poor response to therapy is critical for selecting potential candidates to receive alternative therapeutic approaches that may work better. This is particularly important in patients with relapsing MS, given their partial response to IFN therapy, and the recent advent of more effective therapies; thus, the need to make prompt treatment decisions concerning patients having suboptimal treatment responses, to avoid that MS progresses to such an extent that any treatment adjustment would no longer be effective. In this context, recent studies identify a number of reliable markers of response to IFN therapy.\(^3\)–\(^5\),\(^8\)–\(^10\) Among those studies, the score system proposed by Rio and colleagues\(^4\) has the advantage of combining commonly used, short-term (i.e. 1-year) treatment measures of disease activity, such as clinical relapses, EDSS progression and active MRI lesions, to identify patients who are developing new clinical activity (relapses or disability increases) during the ensuing years.

In the present study, we began by modeling, based on the Rio score, a statistically-optimized scoring system using trial data as a training set, which could easily be used on clinical grounds. This statistical modeling provided us with the best approach for processing the training set, which we then validated using an independent clinical database (the validation set). In both datasets, we found that there was a very close association between the probability of disease progression and any increasing values in the modified Rio score. Indeed, this modified Rio score correctly discriminated the greatest proportion of patients that were falling into the extreme groups, and this useful discrimination was even more evident when the probability of progression was considered since treatment initiation, as it was 20% for patients with score=0 and almost 80% for patients with a score≥2 over a 4-years follow up period. It must be noted that the lowest score (0) included patients with discrete MRI activity during 1 year (i.e. ≤ 4 new T2 lesions), who were thus considered to be at a low risk of disease progression.

Our analysis reported here confirms the difficulty in classifying patients with an intermediate score, who in our study showed a low-to-moderate probability to progress (33% or 43%, depending on 3 or 4 years of assessment of therapy, respectively). This remains an unsolved issue, probably related to the intrinsic difficulty in defining treatment responses in MS. Indeed, there are at least three factors challenging this definition: the high variability within the course of MS, which (especially in the Relapsing–Remitting phase) makes it difficult to attribute a result exclusively to a treatment, even if there is a drastic change in disease activity; the partial efficacy of IFN treatment, which implies the possibility of having responsive patients with residual disease activity; and the definition of disease progression itself, which in MS is not linear through the course of the disease. Despite these limitations in defining treatment responses in MS, the data reported here provide, even in patients in the intermediate score group, a helpful evidence-based quantitative analysis of the likelihood of disease progression after 1 year of IFN treatment, which could help clinicians make treatment decisions. A future step in improving the scoring system will likely focus on the specific group of patients with a score = 1, perhaps by observing their activity for an additional period of time.

In comparison with the previous approaches used to predict treatment response, the new scoring system presented

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3}
\caption{Probability of disability progression from the first year since interferon treatment started (3 years, (a)) and over the follow-up period (4 years, (b)), according to application of the modified Rio score on the “validation set” (Barcelona study data).}
\end{figure}
here has the advantage of being the final result of a statistical modeling procedure. The exclusion of 1-year EDSS progression values from the score is consistent with the observation that a combination of new T2 MRI lesions and relapses seem to be a surrogate for disability progression. Excluding the 1-year EDSS progression also avoids the use of a measure that has high inter-rater variability\(^1\), which would significantly slow down the decision-making process, as it requires at least a 6-month confirmation in a subgroup of patients; however, an EDSS increase in the first few months of therapy should alert us about the possibility of a poor outcome, regardless of the presence of relapses. The cut-offs for MRI lesions found by statistical modeling can be surprisingly higher than those previously proposed;\(^1\) however, in our training set the prognosis of patients with 0–4 new T2 MRI lesions during the first year of therapy resulted very similar. Under these circumstances, the choice of a lower cut-off would have had the effect of lowering the specificity, without a significant gain in sensitivity.

The other novelty provided by the proposed score is the use of the number of new T2 MRI lesions. Previous approaches\(^3,4,8,9\) have used T1-weighted Gd-enhancing lesions to assess MRI activity, since these are easier to detect. Unfortunately, we do not have gadolinium scans for the training set, which precluded us being able to study their role in the model. However, it is known that Gd-enhancing lesions are transient, with a time-course of about 1 month,\(^13\) so that would require very frequent MRI scans to allow for a correct counting of acute lesions. Whereas the classification of a new T2 lesion can be more difficult (i.e. suboptimal MRI scans, incorrect repositioning or presence of confluent lesions) and it is certainly more time consuming, it is also true that T2 signal alterations appearing when a new lesion is formed remain stable in time, as a durable “footprint” of a plaque, thus enabling a correct (and safer) counting of accumulated new lesions, even when the MRI is assessed with much less frequency (i.e. yearly).

In conclusion, the findings reported here translate into clinical practice the observation that the combination of MRI lesions and clinical relapses is a good surrogate for EDSS progression, in IFN-treated patients.\(^6\) This observation is translated into a clinical score that is able to correctly discriminate in a short time the vast majority of patients with RRMS who respond or do not respond to treatment. It must be kept in mind that this specific score is not valid for other DMDs; however, it is likely that similar cohorts of RRMS patients who are treated with IFN or other DMDs will be available for similar analyses soon, making it possible to further validate and refine our proposed score for IFN, as well as to derive new scores for other DMDs, perhaps by including other variables such as immunological markers. This scoring model to determine risk of MS progression early on in IFN treatment is an important step towards the idea of personalized medicine, which offers each patient a treatment that is best suited to him or her, according to his or her risk profile.

**Funding**

This work was partly supported by Merck Serono SA of Geneva, Switzerland (an affiliate of Merck KGaA, Darmstadt, Germany), which allowed for our use of the individual-patient databases. The company collaborated in planning the study and in validating the statistical analyses by double programming.

**Conflict of interest**

Authors’ financial disclosures:

- MPS has received personal compensation for consulting services and for speaking activities from Actelion, Merck Serono, Synthon, Allozyne and Biogen Idec.
- AS, MLS and JR declare no conflict of interest.
- MT has served on scientific advisory boards for Teva Pharmaceutical Industries Ltd., Novartis and Sanofi-Aventis, and has received funding for travel and speaker honoraria from Bayer Schering Pharma, Merck Serono, Teva Pharmaceutical Industries Ltd., Sanofi-Aventis, Biogen Idec and Novartis.
- DL is the Director of the UBC MS/MRI Research Group, which has been contracted to perform central analysis of MRI scans for therapeutic trials by Angiotech, Bayer, Berlex-Schering, Bio-MS, Centocor, Daiichi Sankyo, Genzyme Corp., Hoffmann-LaRoche, Merck Serono, Novartis, Schering-Plough, Teva Neurosciences, Sanofi-Aventis, and Transition Therapeutics. He has received honoraria for speaking activities from Biogen, Merck Serono and Teva Neurosciences.
- PC and BS are employees of Merck Serono SA, Geneva, Switzerland, which is an affiliate of Merck KGaA, Darmstadt, Germany.
- XM has received personal compensation for participating as a member of a company advisory board for: Merck-Serono, Novartis, Teva and Bayer; and for participation in a company-sponsored speaker’s bureau for: Biogen, Bayer, Merck-Serono, Sanofi-Aventis, Novartis and Almirall.
- NDS has received honoraria from Schering, Biogen-Dompè, Teva and Merck Serono for consulting services, speaking and travel support. He serves on advisory boards for Merck Serono. He has received research grant support from the Italian MS Society.

**References**


