Primary progressive multiple sclerosis: a 5-year clinical and MR study

G. T. Ingle,1 V. L. Stevenson,2 D. H. Miller1 and A. J. Thompson1

1Institute of Neurology, University College, London and 2The National Hospital for Neurology and Neurosurgery, London, UK

Correspondence to: Professor Alan J. Thompson, Institute of Neurology, University College, London, Queen Square, London WC1N 3BG, UK
E-mail: a.thompson@ion.ucl.ac.uk

Summary
Longitudinal imaging studies of primary progressive multiple sclerosis (PPMS) have shown significant changes in MR measures over 1 to 2 years. Correlation with clinical change over the same period has not been evident; we investigated the possibility that this is because the period of observation was insufficient for these associations to become apparent. Forty-one patients with PPMS were followed prospectively for 5 years. Patients had clinical [Expanded Disability Status Scale (EDSS) and Multiple Sclerosis Functional Composite Measure (MSFC)] and MRI assessment (brain and spinal cord) at baseline, 1, 2 and 5 years. At 5 years, significant deterioration was seen in all clinical and MRI measures ($P < 0.01$, $P < 0.001$ respectively). Associations were seen between increase in EDSS score and decrease in cord area ($r = 0.31$, $P < 0.05$) and between increase in MSFC and both rate of ventricular enlargement ($r = 0.31$, $P < 0.05$) and increase in $T_2$ load ($r = 0.31$, $P < 0.05$). The rates of change of MR measures were not associated with age or disease duration and were more consistent within than between patients. Longer duration of follow-up demonstrates modest associations between change in clinical and MR measures and provides new insights into the pattern of change within and between individuals with PPMS.

Keywords: atrophy; MRI; multiple sclerosis; primary progressive

Abbreviations: CV = coefficient of variation; EDSS = Expanded Disability Status Scale; FS = functional score; Gd-DTPA = gadolinium diethylene-triamine pentaacetic acid; MSFC = Multiple Sclerosis Functional Composite Measure; PASAT = Paced Auditory Serial Addition Test; PD = proton density; PPMS = primary progressive multiple sclerosis

Introduction
Primary progressive multiple sclerosis (PPMS) is the term applied to patients with multiple sclerosis whose clinical course is progressive from onset without an initial period of clinical relapse and remission (Thompson et al., 1997; McDonnell and Hawkins, 2002). Patients with PPMS have distinctive clinical features, including an older age at onset, a more equal sex incidence and a tendency to present with motor symptoms, typically a progressive spastic paraparesis (Thompson et al., 1997). Possible differences have also been reported between PPMS and relapsing multiple sclerosis in genetics and immunopathology (Neuhaus and Hartung, 2001; Bruck et al., 2002; Kantarci et al., 2002). PPMS is associated with a poor prognosis, which at one time was considered to apply equally to all patients. More recently, differences in outcome associated with the number of systems involved at onset have been reported in a large cohort of patients from Ontario (Cottrell et al., 1999b). The mechanisms underlying progressive disability in PPMS remain incompletely understood and there is, as yet, no treatment that has been convincingly shown to slow its course.

Evidence that major differences in disease mechanisms might exist between PPMS and relapsing multiple sclerosis came from MRI studies where patients with PPMS were found to have less focal abnormality (Thompson et al., 1990, 1991; Filippi et al., 1995). Specifically, lesion loads were lower and there was a reduced frequency of both new lesion formation and of enhancement with the MRI contrast agent gadolinium diethylene-triamine pentaacetic acid (Gd-DTPA). These early findings have been repeated in subsequent studies (Stevenson et al., 1999; Wolinsky et al., 2001). Newer MR techniques have also detected abnormalities in normal appearing tissues, for example on diffusion weighted imaging (Filippi et al., 2001), magnetic transfer imaging (Leary et al., 1999b; Tortorella et al., 2000; Miller et al., 2001) and magnetic resonance spectroscopy (Davie et al., 1997; Leary et al., 1999a; Cucurella et al., 2000). In addition, a network of
increased activation in response to functional MRI (Rocca et al., 2002).

A key issue in relation to all of these techniques is the insight they provide into the mechanisms underlying clinical disability. The relationship between lesion load and clinical disability in PPMS is poor (Stevenson et al., 1999) in both cross-sectional (van Walderveen et al., 2001) and longitudinal studies, despite the latter showing measurable change in lesion load over periods as short as 1 year (Stevenson et al., 2000; Ingle et al., 2002). However, if these changes on MRI are to be used to monitor treatment response in PPMS (Wolinsky et al., 2001), it is essential to demonstrate that they are clinically meaningful. One limitation of existing longitudinal studies in PPMS is that the period of follow-up has been short (2 years maximum) and may be insufficient to detect clinical deterioration given that progression occurs over many decades. Longer periods of follow-up might be necessary to detect MR–clinical correlation. In addition, observation of the pattern of MR changes over an extended time period may provide information about the dynamics of underlying pathological processes. The purpose of this study was therefore to examine the relation of clinical and MR change over a longer period of 5 years, and to evaluate the dynamics of MR change over that time using a range of brain and spinal cord measures.

**Material and methods**

**Patients and evaluation**

Fifty-nine patients with PPMS who had participated in an earlier 2-year MRI study (Stevenson et al., 1999, 2000; Ingle et al., 2002) were asked to return for a further clinical and MR examination at year 5. These patients had originally been selected from patients attending the National Hospital for Neurology and Neurosurgery, Queen Square, London, UK. At 5 years, four patients had died and one patient could not be contacted. Thirteen patients were unable to re-attend because of increased disability but 41 of 54 patients returned for clinical and MRI reassessment at year 5. Of these 41 patients, 15 took part in a phase two trial of interferon-β1a during years 2 and 3, and nine were in the active treatment arm (Leary et al., 2003).

The following clinical measures were obtained at single visits at baseline, years 1, 2 and 5: Expanded Disability Status Scale (EDSS) (Kurtzke, 1983), timed 10 m walk, nine hole peg test (Goodkin et al., 1988) and Paced Auditory Serial Addition Test, 3 s version (PASAT 3) (Gronwall, 1977). The latter three measures were combined to form the Multiple Sclerosis Functional Composite (MSFC) measure. This is a multidimensional measure, reflecting the varied clinical expression of multiple sclerosis across patients and over time. As the three component variables measure different entities (time for timed walk and nine hole peg test, number of correct answers for PASAT 3), a Z score was used to provide a common metric. The Z score for an item indicates how far and in what direction that item deviates from its distribution’s mean, expressed in units of its distribution’s SD (Cutter et al., 1999). Clinical examinations at baseline, year 1 and year 2 were performed by V.L.S. and at year 5 by G.T.I. At the time of the year 5 assessments G.T.I. was blind to all MRI measures. Clinical outcome data on the 13 patients who were unable to re-attend were obtained by means of a structured telephone interview, which allowed the EDSS score to be calculated (Lechner-Scott et al., 2003). In order to assess the reliability of the telephone interview, eight patients had both phone EDSS and standard EDSS and the two measures were correlated highly (r = 0.92), as reported previously (Lechner-Scott et al., 2003).

**MRI acquisition and analysis**

Imaging of brain and spinal cord was carried out at each of the time points with quantification of total brain lesion load, partial brain volume, ventricular volume, and cervical cord area. The same scanner (Signa 1.5T system; General Electric, Milwaukee, WI, USA) was used throughout the study. Each patient underwent T1- and T2/PD (proton density)-weighted spin echo imaging of the brain [T2/PD: repetition time (TR) = 3000 ms, echo time (TE) = 15/90 ms; T1; TR = 600 ms, TE = 20 ms]. All sequences were acquired as contiguous, 3-mm thick axial slices (44 images in total). In the spinal cord, nine contiguous, 3-mm, sagittal T2-weighted (TR = 2500 ms, TE = 90 ms) and T2/PD-weighted (TR = 2500 ms, TE = 45 ms) slices were obtained. A volume acquired inversion prepared gradient echo acquisition of the spinal cord (60 1-mm slices, TR = 15.6 ms, TE = 4.2 ms, inversion time (TI) = 450 ms, FA 20°, matrix 256 × 256) was also performed (fast-spoiled gradient echo) and from the dataset, a series of five contiguous 3-mm axial slices (perpendicular to the spinal cord) were reformatted using the centre of the C2/C3 disc as the caudal land mark. As at baseline, year 1 and year 2 (Stevenson et al., 1999, 2000; Ingle et al., 2002) lesions were marked on hard copy PD-weighted images for all subjects at year 5 with reference to T2-weighted images by a single rater (D.H.M.) who was blind to all clinical measures. These films were then used to calculate the number of new lesions over 5 years (G.T.I.). Individual rates of increase of T2 and T1 load were calculated by linear regression of the four time points for each patient. Cerebral lesion and whole brain acquisition and analysis

Gradient echo acquisition of the spinal cord (60 1-mm slices, TR = 15.6 ms, TE = 4.2 ms, inversion time (TI) = 450 ms, FA 20°, matrix 256 × 256) was also performed (fast-spoiled gradient echo) and from the dataset, a series of five contiguous 3-mm axial slices (perpendicular to the spinal cord) were reformatted using the centre of the C2/C3 disc as the caudal land mark. As at baseline, year 1 and year 2 (Stevenson et al., 1999, 2000; Ingle et al., 2002) lesions were marked on hard copy PD-weighted images for all subjects at year 5 with reference to T2-weighted images by a single rater (D.H.M.) who was blind to all clinical measures. These films were then used to calculate the number of new lesions over 5 years (G.T.I.). Individual rates of increase of T2 and T1 load were calculated by linear regression of the four time points for each patient. Cerebral lesion and whole brain analysis was carried out at a SUN workstation using a local thresholding technique with manual editing. T2 and T1 hypointensity volume were calculated with reference to the marked hard copies. Baseline, and year 1, 2 and 5 images were analysed by a single rater (G.T.I.).

The measures of partial brain volume and cross sectional spinal cord area reflecting atrophy were acquired using previously described methods (Losseff et al., 1996a, b). Brain volume was measured by extracting the brain images from the skull by means of a computer algorithm and by measuring the volume occupied by six 3-mm slices the most caudal at the level of the velum interpositum cerebri. Ventricular
volumes were obtained by analysing the $T_1$-weighted images using an interactive image analysis package (MIDAS) (Brex et al., 2000). Whole-brain regions were obtained using a semi-automated iterative morphological technique originally developed for three-dimensional volumetric scans. Mean signal intensity over these brain regions was calculated. Ventricular regions were outlined using a thresholding technique with the ventricular–brain boundary being set at 60% of the whole brain signal intensity. The ventricular region consisted of the lateral ventricles including the temporal horn, but excluding the third and fourth ventricles. High-signal structures within the ventricles, e.g. blood vessels, were excluded. Ventricular volumes were automatically calculated from the outlined regions by multiplying total area outlined by slice thickness.

The rate of change for each MR measure was found by averaging individual rates of change calculated by linear regression of the four time points for each patient. The effect of subdividing the group into clinically stable and clinically worsening groups was explored by logarithmic transformation of the data (to account for the wide variation in baseline worsening groups) was explored by logarithmic transformation of the data (to account for the wide variation in baseline variation (CV) was calculated for each measure by dividing the SD by the mean.

Measurement reproducibility was assessed for brain lesion load, cerebral atrophy, cross-sectional cord area and ventricular volume by repeating measurements on a dataset of 10 random subjects at an interval of 2 months. The coefficient of variation (CV) was calculated for each measure by dividing the SD by the mean.

**Statistics**

Differences between MR and clinical measures over time were assessed by means of the Wilcoxon signed rank test and correlations were assessed using Spearman’s rank correlation coefficient. The principal comparisons in the study were between change in each of five MR measures ($T_1$ hypointensity, $T_2$ load, partial brain volume, ventricular volume and cord area) and each of two disability measures (MSFC and EDSS). In view of the limited number of comparisons of the principal measures (10) and because of inter-relation of the MR measures ($T_1$ and $T_2$ load, partial brain volume and ventricular volume), it was felt that a correction for multiple comparisons was not appropriate. A larger number of comparisons was made with baseline MR, clinical and demographic measures, and here it was also thought that it would be inappropriate to correct for multiple comparisons as the purpose of this analysis was to identify potentially confounding factors, should they be present. As multiple comparisons are being carried out, the findings should be regarded as exploratory. Variability in the dynamics and consistency of MR behaviour over time between and within patients was assessed using a random intercepts regression model. The trajectories for the five MRI variables were checked for curvature over the time-frame of the study with the MRI variable as response and linear terms in time as predictor.

**Results**

**Baseline**

The cohort was 54% male and had a mean age of 55.7 years ($SD = 10.9$) at 5 years follow-up. The mean disease duration was 16.3 years ($SD = 6.8$) and the mean time between initial and final assessment was 58.4 months ($SD = 5.2$). The 13 patients who were unable to attend at year 5 had greater disability at baseline than those who did return, and had a baseline median EDSS score of 7.0 compared with 6.0 ($P = 0.001$), slower timed walk ($Z$ score of 7.7 compared with 3.7, $P = 0.006$) and poorer MSFC ($-3.5$ compared with $-1.8$, $P = 0.006$). Age, male : female ratio and disease duration were comparable in the two groups. The four patients who died did not differ from the surviving group in their baseline measures or demographics. The 15 patients who had participated in the 2-year trial of interferon-$eta$1a (nine in the active arm) were not found to differ from other patients in the cohort on any demographic, clinical or MRI measure, at baseline or subsequently. This was also the case when the nine patients who had participated in the active arm of the trial were considered separately.

**Clinical change**

The median EDSS score at 5 years for the study group was 6.5, which represented a median increase of 0.5 points (range $-1$ to 6.5) (Fig. 1, Table 1). Change in EDSS score over 5 years is shown in Fig. 2F. For non-attending patients, the median EDSS score by structured telephone interview was 7.5 ($P < 0.001$). There was no difference in the median change in EDSS score between the attending and non-attending groups. Patients with the greatest change in EDSS score had an earlier age of disease onset ($r = -0.24, P = 0.044$) and higher baseline EDSS score ($r = 0.48$, $P < 0.001$), but there was no relation to disease duration. The greatest changes in the functional scores (FS) of the EDSS were in the bowel-bladder, pyramidal and cerebellar components. The MSFC worsened over 5 years by 0.44 points ($SD = 1.1$, $P < 0.05$). Of its components, the mean changes in $Z$ scores were as follows: timed walk $0.97$ ($SD = 2.9$, not significant), nine hole peg test $0.67$ ($SD = 0.88$, $P < 0.001$) and PASAT $-0.32$ ($SD = 1.1$, $P < 0.05$) (see Table 1). At baseline, only a single baseline PASAT was performed and a learning effect may have had an impact on the changes observed.

**MRI measures**

The mean CV for brain lesion load analysis was 3.1% ($SD = 9.6$), and the more automated measures of brain atrophy, cord cross sectional area and ventricular volume produced CVs of 0.17, 0.51 and 0.26%, respectively. The mean rate of change of $T_2$ load was 1.0 ml per year ($SD = 0.92$) and of $T_1$ hypointensity load was 0.4 ml per year ($SD = 0.51$) (Table 2). The individual variation in 5 year change in $T_2$ and $T_1$ hypointensity load by subject with respect to disease onset is
shown in Fig. 2A and B, respectively. Variability in the dynamics and consistency of MR behaviour over time between and within patients was investigated using a random intercepts regression model. In each case, only a small proportion of total variation was due to within patient variation for T2 load (2.5%), T1 hypointensity load (2.9%), ventricular volume (4.2%), partial brain volume (3.1%) and cervical cord volume (4.1%).

The ratio of T2 to T1 load changed during the study from 15.0 (SD = 32.7) at baseline to 8.5 (SD = 14.5) at year 5 ($P = 0.02$) due to a higher percentage rise in T1 (see Table 2). By year 5, all measures were significantly different from baseline (Table 2). The mean number of new brain lesion seen over 5 years was 4.67 (SD = 3.87), while the mean number of new cord lesions was only 0.35 (SD = 0.63).

**Relation between clinical and MRI measures**

Increasing disability was associated with changes in T2 load, ventricular volume and cord area. Deterioration in MSFC was associated with increase in total T2 load ($r = 0.31$, $P = 0.038$) and increase in ventricular volume ($r = 0.31$, $P = 0.044$). Change in cord area was the only measure that was associated with change in EDSS score ($r = 0.31$, $P = 0.038$). Of MSFC components, worsening timed walk ($r = 0.34$, $P = 0.022$) was significantly associated with change in T2 load and change in PASAT score was associated with decrease in brain volume ($r = 0.31$, $P = 0.041$) and increase in ventricular volume ($r = 0.36$, $P = 0.020$). No clinical measure was found to change in association with increasing T1 load (see Table 3). No difference in rates was found when the cohort was subdivided into a clinically stable (no change in EDSS score; $n = 14$) and clinically deteriorating (worsening EDSS score of 0.5 point or more; $n = 27$) groups.

Higher rates of T2 load increase were associated with younger age at onset ($r = 0.30$, $P = 0.047$), shorter disease duration ($r = 0.30$, $P = 0.042$), higher baseline EDSS score ($r = -0.35$, $P = 0.019$) and higher cerebellar FS ($r = 0.37$, $P = 0.013$). Patients with greater increases of T1 hypointensity load had higher baseline cerebellar and brainstem FS ($r = 0.54$ and $P < 0.001$, $r = 0.41$ and $P = 0.005$, respectively) and lower baseline PASAT score ($r = 0.37$, $P = 0.013$), but there was no association with the EDSS score.

Greater rates of reduction in brain volume were associated with longer disease duration ($r = 0.30$, $P = 0.044$), higher baseline cerebellar FS ($r = 0.30$, $P = 0.043$), higher baseline EDSS score ($r = 0.43$, $P = 0.002$), slower baseline timed walk score ($r = 0.34$, $P = 0.021$) and higher baseline MSFC ($r = -0.34$, $P = 0.021$). Higher rates of ventricular enlargement were associated with higher cerebellar and brainstem FS at baseline ($r = 0.46$ and $r = 0.63$, $P = 0.002$ and $P < 0.001$).

**Table 1** Median EDSS score and range, mean and SD for MSFC and MSFC components [timed 10 m walk (TTMW), nine hole peg test (9-HPT) and PASAT 3]

<table>
<thead>
<tr>
<th>EDSS score</th>
<th>Median</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
<th>Mean Z score</th>
<th>SD</th>
<th>Mean Z score</th>
<th>SD</th>
<th>Mean Z score</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>6.0</td>
<td>2.0–8.5</td>
<td>−1.9</td>
<td>2.3</td>
<td>3.9</td>
<td>5.9</td>
<td>−1.0</td>
<td>1.6</td>
<td>−0.8</td>
<td>1.4</td>
</tr>
<tr>
<td>1 year</td>
<td>6.0</td>
<td>2.0–8.5</td>
<td>−1.9</td>
<td>2.5</td>
<td>3.9</td>
<td>6.0</td>
<td>−1.1</td>
<td>1.8</td>
<td>−0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>2 years</td>
<td>6.0</td>
<td>2.0–9.0</td>
<td>−2.1</td>
<td>2.4</td>
<td>4.6</td>
<td>6.1</td>
<td>−1.2</td>
<td>1.8</td>
<td>−0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>5 years</td>
<td>6.5**</td>
<td>3.0–9.0</td>
<td>−2.3*</td>
<td>2.5</td>
<td>4.9</td>
<td>5.9</td>
<td>−1.6**</td>
<td>1.7</td>
<td>−0.5*</td>
<td>1.4</td>
</tr>
</tbody>
</table>

**Significantly different from baseline at a level of $P < 0.001$; *significantly different from baseline at a level of $P < 0.05$, Wilcoxon signed rank test.
respectively) and also with higher PASAT and peg test scores ($r = 0.51$ and $r = 0.33$, $P < 0.001$ and $P = 0.032$, respectively). Greater rate of reduction in cord area was also associated with younger age ($r = 0.42$, $P = 0.004$) and higher baseline sensory FS ($r = 0.31$, $P = 0.039$).

No association was found between new lesion number over 5 years and any clinical measure.

**Relation between MR measures**

Cross-sectional associations between $T_2$ load, $T_1$ hypointensity load, six-slice brain volume and cord area have already been reported for the original cohort at baseline (Stevenson et al., 1999). Larger ventricular volume was associated cross-sectionally with male sex ($r = 0.36$, $P = 0.015$), higher cerebellar FS ($r = 0.31$, $P = 0.040$) and poorer performance on...
PASAT \( (r = 0.34, P = 0.023) \). The cross sectional MR associations of larger ventricular size were with high T2 load \( (r = 0.48, P < 0.001) \), high T1 hypointensity load \( (r = 0.51, P < 0.001) \) and smaller partial brain volume \( (r = 0.46, P = 0.001) \).

The strongest association between changes in MR measures was between T1 hypointensity and T2 lesion load \( (r = 0.7, P < 0.001) \) (Table 4). Changes in cord volumes appeared to be independent of other MR parameters. Ventricular volume change correlated with partial brain volume change but did not correlate with change in any other MR parameter. The association between baseline MR measures and subsequent MR change is shown in Table 5. Baseline T2 load is the measure that shows most correlation with subsequent change in other MR measures, with the exception of change in cord volume.

**Discussion**

This 5-year study demonstrates for the first time a relationship between clinical deterioration and change in MR measures in PPMS. Three important questions arise: how strong is this relationship, is it clinically meaningful, and what are the implications, if any, for disease monitoring in patients with PPMS? Before discussing these points in detail a number of important limitations regarding the interpretation of this study have to be borne in mind. Patients were recruited from a hospital population and were unselected for disease duration. Despite our best efforts, follow-up was incomplete and non-attending patients differed from attending patients in having higher disability at both baseline and 5 years. This might limit the ability of the study to detect MRI changes associated with higher disability. Furthermore, the absence of a control group limits the ability of the study to determine whether subtle changes in MR measures are due to disease effects or simple aging.

In this study there were three MR changes that correlated with clinical change: change in T2 load correlated with change in MSFC \( (r = 0.31) \), change in cord area correlated with change in EDSS score \( (r = 0.31) \) and change in brain atrophy correlated with change in PASAT score \( (r = 0.31 \) and \( r = 0.36) \). No relationship was found with change in T1 load, despite the largest relative changes over 5 years being seen in this measure. The modest correlations between clinical and MR change are not surprising. Most clinical scales in multiple sclerosis are relatively insensitive to clinical change (Hobart et al., 2000) and conventional MR measures have limited histopathological specificity (Barkhof and van Walderveen, 1999). Similar changes in MR appearances could therefore be produced by different pathological processes. Furthermore, there may not be a direct temporal relationship between MR change and clinical change. Pathological changes occurring at one stage of the condition (with resulting changes in MR appearances) might conceivably have a clinical impact that only becomes apparent at a later stage. As such, there may be periods when changes in MR would not associate strongly with clinical change (despite a true relationship). In this
regard the importance of early MR changes has been highlighted in relapsing multiple sclerosis, where $T_2$ load in the earliest clinical stages of the disease correlated with long-term clinical evolution (Brex et al., 2002).

How do these findings build on the results of earlier MR studies of PPMS? Previous studies showed that one of the characteristics of PPMS was a lower rate of new lesion development and enhancement in comparison to relapse onset multiple sclerosis (Thompson et al., 1991; Filippi et al., 1995). For this reason it was initially not clear whether conventional MR measures would show sufficient change over time to render them useful for the purpose of monitoring treatment. It was subsequently shown that measurable changes in several MR parameters could be demonstrated over periods of time as short as 1 year (Stevenson et al., 2000) with, in the case of $T_2$ load, 92% of this change resulting from the enlargement of pre-existing lesions (Stevenson et al., 2002). When patients with PPMS were followed for 1 or 2 years, change in MR parameters was not shown to correlate with clinical change on any measure (Stevenson et al., 2000; Ingle et al., 2002). However, until now there has been little evidence that these changes are clinically meaningful, i.e. associated with clinical change. At 5 years, correlations become apparent and it is likely that the significant factor in the latter case is the longer duration of follow-up, which allowed a greater degree of clinical progression to occur. Specifically, whereas at 1 year only 25% of patients had shown a significant deterioration in EDSS score (Stevenson et al., 2000), at 5 years this figure rose to 65%. These figures are consistent with the probabilities of progression for PPMS as calculated for patients in the London, Ontario, natural history cohort (Cottrell et al., 1999).

This study provides new information on the dynamics of MR change in patients with PPMS, and illustrates how these vary both between and within individuals over time. PPMS is considered to have a universally poor clinical outcome, but

### Table 3 Associations between change in clinical and MR measures

<table>
<thead>
<tr>
<th></th>
<th>EDSS score</th>
<th>MSFC</th>
<th>TTMW</th>
<th>9-HPT</th>
<th>PASAT 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$</td>
<td>$P$</td>
<td>$r$</td>
<td>$P$</td>
<td>$r$</td>
</tr>
<tr>
<td>$T_2$ load</td>
<td>0.22</td>
<td>0.145</td>
<td>0.31*</td>
<td>0.038</td>
<td>-0.34*</td>
</tr>
<tr>
<td>$T_1$ hypointensity</td>
<td>0.22</td>
<td>0.153</td>
<td>0.05</td>
<td>0.740</td>
<td>-1.11</td>
</tr>
<tr>
<td>Ventricular volume</td>
<td>0.19</td>
<td>0.235</td>
<td>0.31*</td>
<td>0.044</td>
<td>0.150</td>
</tr>
<tr>
<td>Brain volume</td>
<td>0.12</td>
<td>0.405</td>
<td>0.31*</td>
<td>0.041</td>
<td>0.245</td>
</tr>
<tr>
<td>Cord area</td>
<td>-0.31*</td>
<td>0.038</td>
<td>-0.03</td>
<td>0.845</td>
<td>-0.150</td>
</tr>
</tbody>
</table>

TTMW = timed 10 m walk; 9-HPT = nine hole peg test. *Significantly different from baseline at a level of $P < 0.05$ using the Wilcoxon signed rank test.

### Table 4 Associations between change in MR measures over 5 years

<table>
<thead>
<tr>
<th></th>
<th>$\Delta T_1$ hypointensity</th>
<th>$\Delta$ Partial brain volume</th>
<th>$\Delta$ Ventricular volume</th>
<th>$\Delta$ Cord area</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta T_2$ load</td>
<td>$r = 0.668, P = 0.000^{**}$</td>
<td>$r = 0.239, P = 0.017^{*}$</td>
<td>$r = 0.183, P = 0.251$</td>
<td>$r = -0.125, P = 0.420$</td>
</tr>
<tr>
<td>$\Delta T_1$ hypointensity</td>
<td>$r = 0.194, P = 0.006^{**}$</td>
<td>$r = 0.259, P = 0.696$</td>
<td>$r = 0.217, P = 0.173$</td>
<td>$r = -0.091, P = 0.556$</td>
</tr>
<tr>
<td>$\Delta$ Partial brain volume</td>
<td>$r = 0.207, P = 0.002^{**}$</td>
<td>$r = 0.663, P = 0.000^{**}$</td>
<td>$r = 0.108, P = 0.465$</td>
<td>$r = 0.108, P = 0.967$</td>
</tr>
<tr>
<td>$\Delta$ Ventricular volume</td>
<td>$r = 0.054, P = 0.183$</td>
<td>$r = 0.054, P = 0.183$</td>
<td>$r = 0.060, P = 0.040$</td>
<td>$r = 0.060, P = 0.207$</td>
</tr>
<tr>
<td>$\Delta$ Cord area</td>
<td>$r = 0.292, P = 0.181$</td>
<td>$r = 0.054, P = 0.181$</td>
<td>$r = 0.054, P = 0.181$</td>
<td>$r = 0.054, P = 0.181$</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level; *correlation is significant at the 0.05 level; $\Delta$ = change in parameters.

### Table 5 Associations between change in MR measures over 5 years and baseline MR measure

<table>
<thead>
<tr>
<th></th>
<th>$\Delta T_2$ load</th>
<th>$\Delta T_1$ hypointensity</th>
<th>$\Delta$ Partial brain volume</th>
<th>$\Delta$ Ventricular volume</th>
<th>$\Delta$ Cord area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline $T_2$ load</td>
<td>$r = 0.410, P = 0.006^{**}$</td>
<td>$r = 0.529, P = 0.000^{**}$</td>
<td>$r = 0.391, P = 0.060$</td>
<td>$r = 0.508, P = 0.040$</td>
<td>$r = 0.040$</td>
</tr>
<tr>
<td>Baseline $T_1$ hypointensity</td>
<td>$r = 0.194, P = 0.000^{**}$</td>
<td>$r = 0.451, P = 0.000^{**}$</td>
<td>$r = 0.464, P = 0.015$</td>
<td>$r = 0.687, P = 0.037$</td>
<td>$r = 0.967$</td>
</tr>
<tr>
<td>Baseline partial brain volume</td>
<td>$r = 0.197, P = 0.000^{**}$</td>
<td>$r = 0.285, P = 0.000^{**}$</td>
<td>$r = 0.204, P = 0.345$</td>
<td>$r = 0.345, P = 0.065$</td>
<td>$r = 0.521$</td>
</tr>
<tr>
<td>Baseline ventricular volume</td>
<td>$r = 0.150, P = 0.000^{**}$</td>
<td>$r = 0.207, P = 0.000^{**}$</td>
<td>$r = 0.162, P = 0.020^{*}$</td>
<td>$r = 0.345, P = 0.065$</td>
<td>$r = 0.697$</td>
</tr>
<tr>
<td>Baseline cord area</td>
<td>$r = 0.292, P = 0.000^{**}$</td>
<td>$r = 0.206, P = 0.000^{**}$</td>
<td>$r = 0.023, P = 0.689$</td>
<td>$r = 0.173, P = 0.236$</td>
<td>$r = 0.662$</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level; *correlation is significant at the 0.05 level; $\Delta$ = change in parameters.
the rate of clinical deterioration in this cohort over 5 years varied greatly. This variability was also seen in the MR measures, as shown in Fig. 2, with some patients showing large amounts of change over the study period and others much less. In contrast, the rate of change in individual patients appears relatively constant over time. This relative consistency of both clinical and MR behaviour in PPMS within patients over time with variability in outcome between patients suggests differences in underlying mechanisms of disease and repair between patients that would also seem to be relatively constant over time. Such constancy raises the possibility that useful prognostic information could be gained from the early assessment of the rate of MR change in PPMS and allow early identification of those patients at risk of poor outcome.

Although both MR and clinical changes may be relatively consistent over time, their relationship may not be straightforward, and it is probably too simplistic to infer that patients with the greatest change on MR will always show the greatest clinical change. In this study, there is a wide variation in rate of increase of T1 hypointensity, but no association with the rate of clinical progression. One could hypothesize that change in some MR measures, or more properly the underlying pathological processes that they imply, may be essentially continuous throughout the course of the disease, but only cause disability at certain critical periods, or after certain threshold points have been reached. At a pathological level it has been suggested that such a threshold effect may operate with respect to axonal loss (Trapp et al., 1999). It can be seen from Fig. 2 that patients can have identical lesion loads at a single point in time and yet differ in terms of their rate of change of MR abnormality. This study suggests that in certain cases it is the change over time of an MR variable, rather than its absolute value, that is the important measure. The long-term predictive value of short-term change measures has, however, yet to be demonstrated.

Treatment trials in PPMS have, to date, used progression of disability as their primary outcome measure. However, as clinical evolution can be slow and its assessment difficult, there has been interest in identifying MRI outcome measures as so-called ‘surrogate markers’. MRI measures have the advantage of being objective, sensitive and quantitative, but to be useful as surrogates it needs to be shown that they accurately reflect differences in clinical behaviour between patients. It has been suggested that such measures need to be validated: the measure should predict future clinical disease, the effect of treatment on clinical disease must be explained by the effect on the surrogate and the surrogate should not be restricted to use with a single treatment, but should ideally respond in a similar and predictable manner to all effective treatments (McFarland et al., 2002). This natural history study supports the view that conventional MR measures do relate to clinical status, and suggests that atrophy measures may provide additional information. While T1 load and brain atrophy measures are somewhat correlated, cord atrophy does not correlate with other MR changes. It therefore appears to be providing independent information, suggesting that cord atrophy measures should be included in MR studies in this group. Alternative cord measures such as quantification of signal abnormality on T2-weighted MRI are poor markers of pathological change (Bergers et al., 2002). Validation of the other criteria will need to await the development of effective treatments in full-scale clinical trials. At present it seems reasonable to expect that treatments which reduce the rate of increase of T2 load, ventricular volume and cord atrophy in the shorter term would, in the longer term, be shown to have clinical benefit.

This study supports the usefulness of longitudinal examination. In the future, it may be appropriate to focus such longitudinal MR studies on the earliest clinical stages of the condition using newer, and pathologically more specific, MR measures if we are to understand the underlying pathological mechanisms that give rise to disability in this condition.

References
Davie CA, Barker GI, Thompson AJ, Tofts PS, McDonald WI, Miller DH. 1H magnetic resonance spectroscopy of chronic cerebral...


