

# Disability outcome measures in therapeutic trials of relapsing-remitting multiple sclerosis: effects of heterogeneity of disease course in placebo cohorts

Clarence Liu, Lance D Blumhardt

## Abstract

**Objectives**—Recent phase III clinical trials of immunomodulatory therapies in relapsing-remitting multiple sclerosis have shown significant benefits of active treatment on relapse related end points, but effects on disability outcomes have been inconsistent. These apparent discrepancies could be due to differences in the clinical end points employed, the behaviour of placebo cohorts, or both.

**Methods**—Disability data from the placebo cohorts of two large phase III studies, the United States glatiramer acetate trial (Copolymer 1 Multiple Sclerosis Study Group) and the multinational interferon  $\beta$ -1a trial (PRISMS Study Group) were combined and masked (n=313). Two groups of disability outcome measures were assessed. Firstly, measures of disability change (2 year EDSS difference and area under the EDSS/time curve, AUC) were calculated. Secondly, conventional disease progression end points (“confirmed progression” and “worsening to EDSS 6.0”) were evaluated by using Kaplan-Meier analysis and compared with a categorical classification based on EDSS trends.

**Results**—The average increase in disability for the entire cohort as assessed by mean 2 year EDSS change (<0.5 EDSS point) or mean AUC (+0.57 EDSS-years) was small. For the “confirmed progression” end points, increasing the stringency of the definition lowered their incidence (from 32% with 1.0 point at 3 months, to 9% with 2.0 points at 6 months), but did not improve the positive predictive accuracy for “sustained progression” maintained to the end of the study. The error rate for this outcome was about 50%. Worsening to EDSS 6.0 was a more reliable end point, but had even lower sensitivity (incidence <10%). EDSS trend analysis showed markedly heterogeneous disease courses, which were then categorised into “stable” (26%), “relapsing-remitting” (59%), and “progressive” (15%) courses. Patients with the last course had deteriorated considerably by the end of 2 years (mean worsening of 2.0 EDSS points).

**Conclusion**—In relapsing-remitting multiple sclerosis treatment trials, the conventional measure of mean EDSS change has low sensitivity, whereas the widely

applied confirmed progression end points have high error rates regardless of their definition stringency. Alternative methods with better data utilisation include AUC summary measures and categorical disease trend analysis. The heterogeneity of disability outcomes in short trials, combined with unreliable clinical end points, diminishes the credibility of therapeutic claims aimed at reducing irreversible neurological deficits. The behaviour of patients treated with placebo should be carefully analysed before conclusions are drawn on the efficacy of putative treatments.

(J Neurol Neurosurg Psychiatry 2000;68:450-457)

Keywords: multiple sclerosis; disability outcome measures; treatment trials

Recent double blind, placebo controlled phase III therapeutic trials in relapsing-remitting multiple sclerosis have consistently demonstrated the efficacy of immunomodulatory drugs for reducing disease activity.<sup>1-5</sup> Less well studied are the severity of transient symptoms and the effects on progression of disability.<sup>6-8</sup> Data from relatively short trials of relapsing-remitting multiple sclerosis are undoubtedly difficult to analyse and interpret, as many clinical exacerbations, although disabling, may be reversible and the accumulation of fixed neurological dysfunction across a representative cohort may take many years.<sup>9-10</sup> Evidence that new therapies prevent *irreversible* worsening of disability is less consistent.<sup>11-12</sup> One example is the contrast between the North American interferon  $\beta$ -1b and intramuscular interferon  $\beta$ -1a studies; the first showed no effects on disability,<sup>13</sup> whereas the second demonstrated apparent treatment benefits in reducing disease progression.<sup>3-14</sup> Although this may be due to genuine differences in the effects of these agents, interpretation is confounded by factors such as variance in the outcome measures employed and fluctuations in the behaviour of the placebo cohorts.

We attempted to address these issues by examining disability data, as measured by the expanded disability status scale (EDSS),<sup>15</sup> from the placebo arms of two randomised, double blind, 2 year therapeutic trials: the United States study of copolymer 1 (or glatiramer acetate, *Copaxone*®), investigated by the Copolymer 1 Multiple Sclerosis Study Group<sup>2</sup> and the multinational study of interferon  $\beta$ -1a (*Rebif*®), investigated by the PRISMS Study

Division of Clinical Neurology,  
Department of Medicine, University Hospital, Queen's Medical Centre, Nottingham, UK  
C Liu  
L D Blumhardt

Correspondence to:  
Professor L D Blumhardt,  
Division of Clinical Neurology, Department of Medicine, University Hospital, Queen's Medical Centre, Nottingham NG7 2UH, UK

Received 20 July 1999 and in final form  
12 November 1999  
Accepted 25 November 1999

Group.<sup>5</sup> Both trials had similar baseline enrolment characteristics and were well conducted, especially with respect to their low dropout rates and adequate treatment blinding. We were thus able to evaluate in-study disability of the patients treated with placebo with two methods which quantify disability change (the “2 year EDSS difference” from trial entry to completion and the summary measure “area under the disability/time curve” (AUC)),<sup>16</sup> and three measures which assess disease progression (“confirmed EDSS progression”, “worsening to EDSS 6.0”, and “categorical EDSS trend analysis”).

## Methods

The enrolment criteria and trial designs of the original phase III studies have been previously published.<sup>2–5</sup> The baseline demographics of the patients in the placebo arms of the two trials were indistinguishable (table 1), enabling all EDSS data to be masked and incorporated into a single database. Due to the non-linearity and varying staying times at different levels on the EDSS scale,<sup>17</sup> a worsening of 0.5 point between EDSS of 5.5 and 7.0 was adjusted to be equivalent to 1.0 point.<sup>18–19</sup>

### DISABILITY CHANGE

#### *Two year EDSS difference*

Conventional EDSS change over the 2 year trial was obtained from the difference between the initial baseline and final (2 year) EDSS scores.

#### *AUC analysis*

A summary measure statistic was derived by integrating the area under the plotted EDSS/time curve (AUC) normalised to entry disability level,<sup>16</sup> to provide an index of in-trial morbidity change. Analysis was performed on data acquired solely from scheduled visits to avoid possible bias due to differential sampling rates between treatment arms and, to some extent, the effects of transient relapse related disability.

### DISEASE PROGRESSION

#### *Confirmed progression*

This end point, which is widely used in clinical trials to assess therapeutic effects, is often interpreted as an indicator of irreversible worsening of disability. We have applied this outcome measure with increasing stringency (1.0 EDSS point increase at 3 months, 1.0 point increase at 6 months, 2.0 point increase at 3 months, and 2.0 point increase at 6 months) to data from scheduled visits. The

Table 1 Baseline characteristics of placebo patients in two phase III trials

|                                  | Copolymer-1<br>(sc) | Interferon $\beta$ -1a<br>(sc) |
|----------------------------------|---------------------|--------------------------------|
| Number of patients               | 126                 | 187                            |
| Mean entry EDSS (SD)             | 2.4 (1.3)           | 2.4 (1.2)                      |
| Mean disease duration (SD)       | 6.6 (5.1)           | 6.1 (4.8)                      |
| Mean entry 2 y relapse rate (SD) | 2.9 (1.1)           | 3.0 (1.3)                      |
| Mean age (SD)                    | 34.3 (6.5)          | 34.7 (7.5)                     |
| Male: female ratio               | 1:3.2               | 1:3                            |
| Participating centres            | 11                  | 22                             |

irreversibility (positive predictive accuracy) of these outcomes was determined by comparisons with sustained progression end points, defined from individual EDSS/time plots. To qualify as sustained, a particular confirmed step worsening had to be maintained to the end of the study. Deteriorations during the final 3 or 6 months, for 3 month or 6 month end points respectively, were also excluded.

#### *Worsening to EDSS 6.0*

This end point, indicating the need for unilateral ambulatory assistance, has been found to be a useful natural history landmark.<sup>9</sup> It was examined with confirmation at 3, 6, 9, and 12 months and tested for positive predictive accuracy by comparisons against sustained worsening to EDSS 6.0.

#### *Categorical EDSS trend analysis*

All complete 2 year EDSS datasets (n=289) were plotted for individual patients. Disease trends could be categorised into six subsets according to the following definitions:

*Minimal change*—Fluctuations of no more than  $\pm 0.5$  EDSS point (the generally accepted range of EDSS interrater variability) from baseline level (or no change if EDSS  $\geq 5.5$ ).<sup>19</sup>

*Erroneous progression*—An increase of at least 1.0 EDSS point confirmed at 3 months, but not sustained to the end of the trial (as defined in the previous section).

*Erroneous improvement*—A decrease of at least 1.0 EDSS point confirmed at 3 months but not sustained to the end of the trial.

*Sustained progression*—An increase of at least 1.0 EDSS point confirmed at 3 months and sustained until the end of the study.

*Sustained improvement*—A decrease of at least 1.0 EDSS point confirmed at 3 months and sustained until the end of the study.

*Fluctuating*—All other EDSS plots with unsustained increases or decreases during the study not meeting the above definitions.

After statistical analyses (see below), these six subsets were collapsed into three clinically meaningful categories of disease courses—namely:

*Stable*—Including the subsets sustained improvement and minimal change.

*Relapsing–remitting*—Including the subsets fluctuating course, erroneous progression, and erroneous improvement.

*Progressive*—Including the subset with sustained progression.

### STATISTICS

End points for disability change were compared with the Mann-Whitney *U* test. Relations between 2 year EDSS difference and AUC were explored with (1) Spearman’s rank correlation, and (2) stratification of the AUC data by 1.0 point steps of EDSS change. The stratified data sets were normally distributed (Shapiro-Wilk *W* test) and multiple comparisons were performed with one way analysis of variance (ANOVA) and Bonferroni multiple *t* tests.

Kaplan-Meier survival analyses were performed for both confirmed and sustained progression outcomes, with censoring of patients

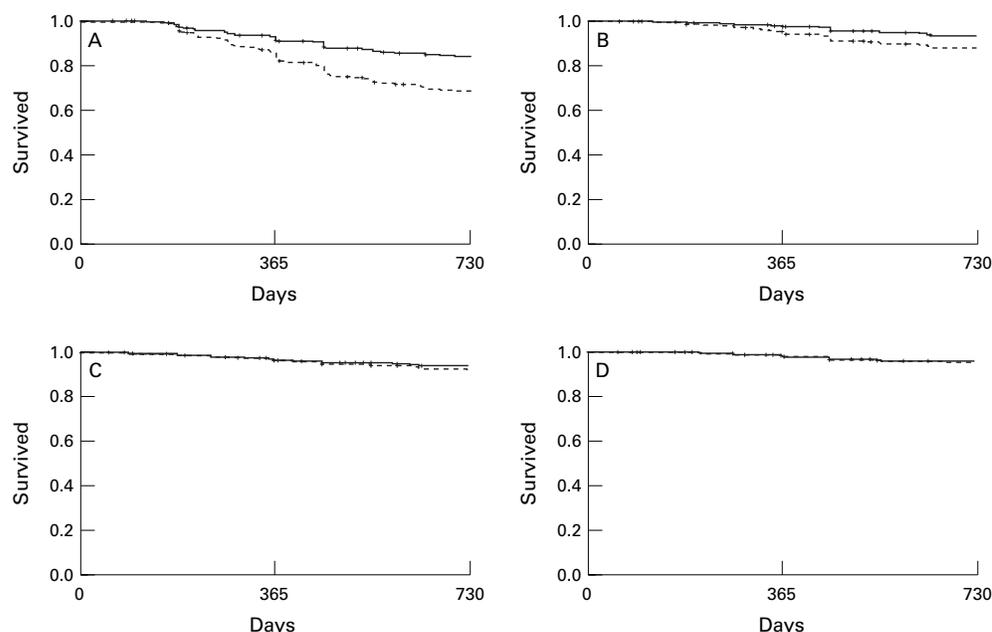


Figure 1 Examples of Kaplan-Meier survival curves with cumulative survival probability plotted *v* days in study, for different EDSS progression end points. Comparison of confirmed progression (dotted) *v* sustained progression (solid) with (A) 1.0 EDSS point progression at 3 months (log rank  $p < 0.0001$ ), and (B) 2.0 EDSS points progression at 3 months (log rank  $p = 0.018$ ). Comparison of confirmed worsening to EDSS 6.0 (dotted) *v* sustained worsening (solid) at (C) 3 months (log rank  $p = 0.265$ ), and (D) 6 months (log rank  $p = 0.562$ ).

who had not reached the particular end point in question, either at study completion, or at the time of dropout. The log rank test was used to measure the amount of irreversibility of confirmed against sustained end points at study completion. The positive predictive value of each confirmed end point was obtained. With categorical EDSS trend analysis, intercategory disability data (non-parametric) were compared using the Kruskal-Wallis test. The association of baseline demographical factors in predicting subsequent disease courses was investigated by logistic regression analysis.

## Results

### DISABILITY CHANGE

For the total cohort, the mean baseline and final EDSS were 2.42 (SD 1.24) and 2.81 (SD 1.76) respectively ( $p < 0.0001$ ). The mean unadjusted 2 year EDSS difference was +0.38 (SD 1.16) (after adjustment for EDSS non-linearity, mean difference was +0.45 (SD

1.32)). The mean change estimated by AUC calculations was +0.57 EDSS-years (SD 1.65).

The two disability change end points, 2 year EDSS difference and AUC, were significantly correlated ( $r = 0.76$ ;  $p < 0.0001$ ). When AUC data were stratified by 1.0 point steps of 2 year EDSS difference, there was increased sensitivity of the AUC for measuring disability change as shown by its greater data spread per EDSS point shift. Multiple comparisons demonstrated that the stratified sets were significantly different from each other (all comparisons,  $p < 0.005$ ), thus maintaining separation by conventional EDSS changes.

### DISEASE PROGRESSION

#### Confirmed progression

With increasingly more demanding outcome definitions employed in the Kaplan-Meier analyses, there was a gradual increase in the cumulative survival probability and a reduction in the number of patients reaching the end points, utilising either confirmed (32% for 1.0 point at 3 months; 9% for 2.0 points at 6

Table 2 Disease progression data for the total cohort ( $n = 313$ ). Comparisons of Kaplan-Meier analyses of “confirmed” *v* “sustained” progression end points, with (A) EDSS point progression and (B) worsening to EDSS 6.0

| (A) EDSS point increase   | Confirmed progression |               |              | Sustained progression |               |              | Log rank <i>p</i> value | Positive prediction of confirmed for sustained |
|---------------------------|-----------------------|---------------|--------------|-----------------------|---------------|--------------|-------------------------|--|
|                           | <i>c</i> (10%)        | <i>S</i> (1y) | <i>n</i> (%) | <i>c</i> (10%)        | <i>S</i> (1y) | <i>n</i> (%) |                         |  |
| +1.0-point 3 months       | 274                   | 0.83          | 101 (32%)    | 456                   | 0.91          | 48 (15%)     | <0.0001                 | 0.48   |
| +1.0-point 6 months       | 457                   | 0.94          | 67 (21%)     | 547                   | 0.95          | 44 (14%)     | 0.024                   | 0.67   |
| +2.0-points 3 months      | 548                   | 0.94          | 38 (12%)     | NR                    | 0.97          | 20 (6%)      | 0.018                   | 0.53   |
| +2.0-points 6 months      | NR                    | 0.98          | 29 (9%)      | NR                    | 0.98          | 16 (5%)      | 0.084                   | 0.55   |
|                           | Confirmed progression |               |              | Sustained progression |               |              |                         |  |
| (B) Worsening to EDSS 6.0 | <i>c</i> (5%)         | <i>S</i> (1y) | <i>n</i> (%) | <i>c</i> (5%)         | <i>S</i> (1y) | <i>n</i> (%) | Log rank <i>p</i> value | Positive prediction of confirmed for sustained |
| 3 months                  | 526                   | 0.96          | 24 (8%)      | 637                   | 0.97          | 17 (5%)      | 0.265                   | 0.71   |
| 6 months                  | 729                   | 0.98          | 17 (5%)      | NR                    | 0.98          | 12 (4%)      | 0.562                   | 0.71   |
| 9 months                  | NR                    | 0.99          | 12 (4%)      | NR                    | 0.99          | 10 (3%)      | 0.669                   | 0.83   |
| 12 months                 | NR                    | 1             | 10 (3%)      | NR                    | 1             | 9 (3%)       | 1                       | 0.90   |

*c*=Centiles (days); *S*(1y)=cumulative survival probability at 1 year; *n*=number (and %) of patients reaching end point; NR=end point not reached.

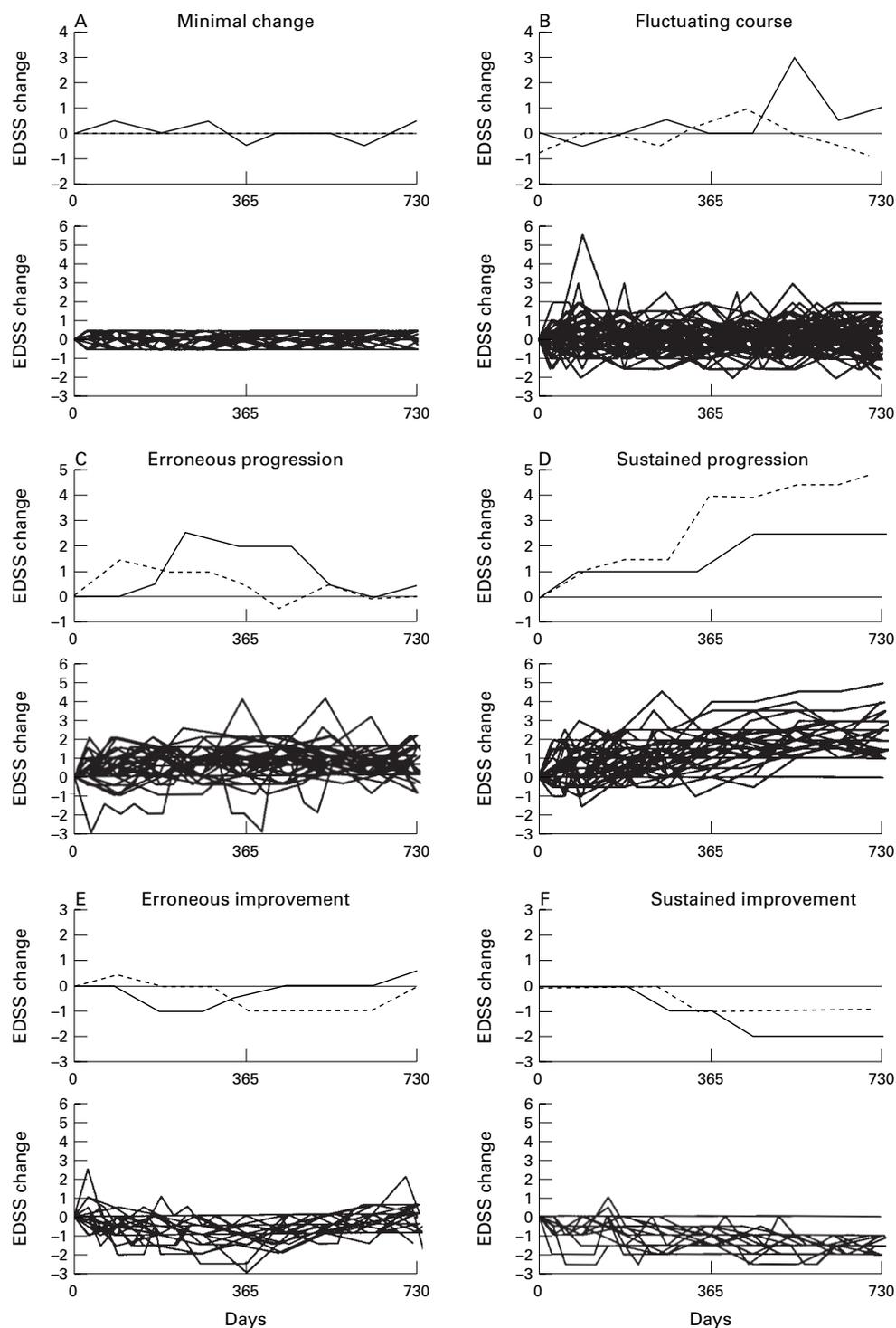


Figure 2 Disability plots from scheduled visits (EDSS change from baseline *v* days in study) of patients with complete datasets ( $n=289$ ) encoded into six subsets: (A) minimal change, (B) fluctuating course, (C) erroneous progression, (D) sustained progression, (E) erroneous improvement, and (F) sustained improvement. In each subset, the upper panel shows two representative examples, the bottom panel shows the total series of EDSS plots from all patients.

months) or sustained progression (15% for 1.0 point at 3 months; 4% for 2.0 points at 6 months). However, regardless of the stringency of such definitions, about 50% of the decisions satisfying confirmed progression were erroneous when compared with sustained progression (positive predictive accuracy of 48% for 1.0 point at 3 months and 55% for 2.0 points at 6 months) (fig 1 A–B, table 2).

#### Worsening to EDSS 6.0

Analyses of confirmed worsening to EDSS 6.0 showed that the error rates for the whole cohort were relatively low (29% and 10% at 3 and 12 months, respectively) and the paired confirmed and sustained progression curves were not significantly different (fig 1 C–D, table 2). On the other hand, the incidence rates of these outcome measures for sustained progression

Table 3 Disease progression data for the total cohort with complete datasets (n=289). In trial disability changes of categorical EDSS trend analyses by (A) six subsets and (B) three disease course categories

|                                     | n (%)    | Two year EDSS difference |                     | AUC (in EDSS-year) |                        |
|-------------------------------------|----------|--------------------------|---------------------|--------------------|------------------------|
|                                     |          | Mean (SD)                | Median (Q1, Q3)     | Mean (SD)          | Median (Q1, Q3)        |
| (A) Six subsets of EDSS trends      |          |                          |                     |                    |                        |
| Minimal change                      | 57 (20)  | +0.03 (0.36)             | 0 (0 to +0.5)       | +0.02 (0.50)       | 0 (-0.35 to +0.25)     |
| Sustained improvement               | 17 (6)   | -1.44 (0.43)             | -1.5 (-2.0 to -1.0) | -2.22 (0.70)       | -2.07 (-2.61 to -1.64) |
| Erroneous improvement               | 24 (8)   | -0.27 (0.86)             | 0 (-1.0 to +0.5)    | -1.30 (0.54)       | -1.31 (-1.63 to -0.88) |
| Fluctuating course                  | 107 (37) | +0.12 (0.85)             | 0 (-0.5 to +0.5)    | +0.21 (0.67)       | +0.27 (+0.26 to +0.72) |
| Erroneous progression               | 40 (14)  | +0.53 (0.79)             | +0.5 (0 to +1.0)    | +1.26 (0.76)       | +1.28 (+0.77 to +1.95) |
| Sustained progression               | 44 (15)  | +2.03 (0.98)             | +2.0 (+1.0 to +2.5) | +3.14 (1.58)       | +2.95 (+1.83 to +4.06) |
| (B) Three disease course categories |          |                          |                     |                    |                        |
| Stable course                       | 74 (26)  | -0.31 (0.73)             | 0 (-0.5 to +0.5)    | -0.50 (1.10)       | -0.06 (-0.82 to +0.12) |
| Relapsing-remitting                 | 171 (59) | +0.16 (0.87)             | 0 (-0.5 to 0)       | +0.24 (1.01)       | +0.33 (-0.51 to +0.94) |
| Progressive course                  | 44 (15)  | +2.03 (0.98)             | +1.5 (+1.0 to +2.5) | +3.14 (1.58)       | +2.95 (+1.83 to +4.06) |

were low (5% at 3 months and 3% at 12 months).

#### Categorical EDSS trend analysis

The individual EDSS plots from 289 patients with complete 2 year data were encoded into the six subsets described above (fig 2). One hundred and seven patients (37%) followed a fluctuating course, whereas 57 (20%) experienced nil, or only minimal change. Forty one patients (14%) initially showed improving in-trial disability trends, 17 of whom maintained improvement to trial end (sustained improvement). Forty out of 84 patients (29%) who originally exhibited progressive trends subsequently improved (erroneous progression). The worsening of disability scores in the remaining 44 patients (15%) was maintained to trial end (sustained progression).

The disability characteristics as measured by 2 year EDSS difference and AUC were calculated for each subset (table 3). Multiple comparisons (Kruskal-Wallis test) between the six subsets demonstrated that distinct segregation was not consistently accomplished, particularly when comparing differences between the subsets classified as erroneous improvement, fluctuating course, and erroneous progression. These subsets were therefore collapsed into three categories of in-trial disease course (stable, relapsing-remitting, and progressive) which were ordered and separable using the above disability end points (table 3). Thus, 59% of patients followed a relapsing-remitting course, 26% a stable course, and 15% a progressive course (figure 3).

Analysis of the baseline demographic data (age, sex ratio, pretrial relapse rate, and disease duration) did not show any intercategory

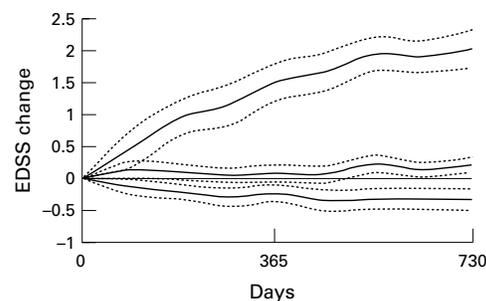


Figure 3 Mean disability courses (EDSS change from baseline (95% CI) of patients with complete datasets (n=289) during the study divided into three categories: bottom curve stable course; middle curve relapsing-remitting course; top curve progressive course.

differences or associations between clinical parameters at study entry and subsequent disease course categories, although higher EDSS at baseline was significantly associated with a progressive course (p=0.01).

#### Discussion

The behaviour of the placebo arm can substantially influence the outcome of a therapeutic trial. This is seen in recent years with the results from the North American phase III interferon  $\beta$ -1b<sup>1</sup> and intramuscular interferon  $\beta$ -1a<sup>3</sup> studies in relapsing-remitting multiple sclerosis. With disease progression defined as confirmed deterioration by 1.0 EDSS point, the interferon  $\beta$ -1b treated arms were not significantly different from placebo,<sup>13</sup> whereas patients receiving interferon  $\beta$ -1a benefited significantly more than their placebo counterparts on this and additional more stringent end points.<sup>3,14</sup> However, it has been pointed out that differences inherent in the disability of the patients in the interferon  $\beta$ -1a placebo arm may have confounded the results: the 1 year confirmed progression rates were similar for patients receiving interferon  $\beta$ -1b, interferon  $\beta$ -1b placebo, and interferon  $\beta$ -1a (11 to 13%), but much worse (22%) for those randomised to interferon  $\beta$ -1a placebo.<sup>11</sup> Hence the behaviour of placebo groups should be examined thoroughly before emphatic conclusions are made about the benefits of putative therapies on disability.

In our present retrospective analysis, data from the placebo arms of two large relapsing-remitting multiple sclerosis treatment trials<sup>5</sup> were combined and evaluated. This was possible as the two cohorts were very similar in their baseline clinical demography and both studies were adequately blinded with low dropout rates. This approach also ensured data masking, as our aim was not to compare the efficacies of different drugs. In-trial relapse rates were not examined, due to the different definitions used in various trial protocols<sup>8</sup> and the established significant benefits of immunomodulatory therapies on relapse reduction. Our analyses focused on the disability changes measured by the EDSS, which have caused much debate among neurologists. Controversies include the difficulties associated with the scale (which are beyond the scope of this article), problems interpreting the conventional end points for progression, and the clinical meaningfulness of such end points in relatively

Table 4 Comparisons with disability data from placebo cohorts of other published phase III trials. End points measured at 2 years unless otherwise stated

|                                       | Entry EDSS range | 2 year EDSS difference (unadjusted) | EDSS confirmed progression end points |                   |                 |
|---------------------------------------|------------------|-------------------------------------|---------------------------------------|-------------------|-----------------|
|                                       |                  |                                     | 1.0-pt 3 months                       | 1.0-pt 6 months   | 2.0-pt 6 months |
| This study:                           |                  |                                     |                                       |                   |                 |
| Total cohort                          | 0 to 5.0         | +0.38                               | 32%                                   | 21%               | 9%              |
| Other relapsing–remitting MS studies: |                  |                                     |                                       |                   |                 |
| INFβ-1b <sup>1</sup>                  | 0 to 5.5         | NA                                  | 28% (3 years)                         | NA                | NA              |
| INFβ-1a <sup>3</sup>                  | 1.0 to 3.5       | +0.74                               | NA                                    | 35%               | 18%             |
| IVIg <sup>4</sup>                     | 1.0 to 6.0       | +0.12                               | NA                                    | 23% (unconfirmed) |                 |
| Secondary progressive MS study:       |                  |                                     |                                       |                   |                 |
| INFβ-1b <sup>31</sup>                 | 3.0 to 6.5       | +0.60 (3 years)                     | 50% (3 years)                         | NA                | NA              |

The 1.0 point 6 months end point for the IVIg trial was *not* confirmed as this was not an outcome measure used in the study.

short studies of relapsing–remitting multiple sclerosis. We discuss in turn the outcomes with respect to in-trial EDSS changes and the determinants of disease progression.

#### IN-TRIAL DISABILITY CHANGES

The mean adjusted 2 year EDSS difference for the entire combined placebo cohort was <+0.5. This small change is not unexpected when contrasted with natural history series<sup>10 18 20</sup> and is comparable with most published phase III trials of relapsing–remitting multiple sclerosis (with the exception of the intramuscular INFβ-1a study) (table 4). In addition, the use of the first and final snapshot assessment scores to derive disability changes in a relapsing and remitting disease ignores the fluctuating in-trial morbidity changes commonly experienced by these patients.

An alternative technique is to employ the summary measure obtained by integrating the area under the EDSS/time curve (AUC)<sup>16</sup> with normalisation to baseline EDSS, to capture each patient's within study disability experience. Our analysis demonstrates that AUC is more sensitive to in-trial disability changes, and allows further calculations using continuous data without destabilising the divisions in regard to EDSS steps. The pros and cons of the AUC technique have been previously discussed.<sup>16</sup> One problem is that it does not reflect disease trends.<sup>21</sup> A patient with a period of improvement followed by deterioration may have the same AUC as one with the opposite temporal sequence, or another whose EDSS does not change throughout the study. Therefore, disease trends need to be separately evaluated.

#### PROGRESSION END POINTS

The behaviour of patients on placebo with regard to confirmed progression end points has been examined systematically in progressive multiple sclerosis.<sup>22</sup> By varying the definitions of whether confirmation or treatment failure was allowable, the proportion of patients reaching the end point could vary by 20%. Another study on design strategies of clinical trials did include 91 relapsing–remitting patients, but the disease courses were defined solely on DSS change per year and the authors' objective was to estimate the power necessary for trial planning.<sup>18</sup> In view of the demonstrated effectiveness of new treatments on disease progression in recently completed and published phase III placebo controlled trials of relapsing–

remitting multiple sclerosis, our emphasis has shifted to the interpretation of such end points. Although success for a drug on a confirmed progression end point must undoubtedly indicate certain therapeutic effects, it is often equated with a beneficial reduction of permanent or sustained disease progression.<sup>23</sup> However, it has also been pointed out that 47% of the patients reaching such a treatment failure end point in the first year of the intramuscular interferon β-1a study subsequently improved.<sup>12</sup>

We tested successively more vigorous definitions on our cohort. The accuracy with which the confirmed progression end point predicted sustained progression was never better than 67%. In addition, as our criteria for sustained progression only considered data up to 2 years, the proportion of erroneous progressors was probably underestimated. Although there is certainly noise in our data, the reversibility of confirmed progression must be partly due to prolonged relapses. Information on the *duration* of exacerbations from natural history series is scarce, although it has been documented that in patients with early multiple sclerosis, 22% of initial episodes last between 3 and 12 months and 10% between 6 and 12 months.<sup>24</sup> On the other hand, the use of worsening to EDSS 6.0 (the requirement for unilateral ambulatory assistance), which has been considered an important milestone in disease progression in multiple sclerosis,<sup>9</sup> disclosed no significant differences between paired confirmed and sustained plots. However, this advantage is offset by the very low incidence of the end point. This is not surprising, given that the median rate of reaching this level is between 15 to 20 years from diagnosis.<sup>10 20 25</sup> Hence, making outcome definitions more stringent to achieve increasing stability and lack of fluctuation is confounded by decreasing sensitivity. This led us to explore other techniques for predicting disability outcome in relatively short studies.

Our initial premises on disease trend analyses were based on the utilisation of as much trial data as possible (all complete serial 2 year EDSS datasets) and the separation of patients whose progression was sustained to trial end from those with recovery from confirmed progression (erroneous progression). The minimal change subset was defined according to the generally accepted range for EDSS interrater reproducibility error<sup>19</sup> while the rest of the cohort typically experienced a fluctuating course. Although all patients fell within the definitions of our six subsets, comparisons of

the in-trial disability change end points showed incomplete subset separation. This prompted a secondary assignment of patients into just three categories which were statistically separable, had diverse EDSS plots (with distinct 95% confidence intervals), and clinical relevance. As expected with this classification, most patients (59%) indeed followed a relapsing-remitting course. Interestingly, only 14% maintained a progression of EDSS worsening throughout the study. These patients deteriorated markedly, with a mean 2 year EDSS difference of over 2.0 points and a mean AUC of +3.14 EDSS-years. With the exception of the entry EDSS, which is a well known risk factor for progression,<sup>17</sup> we found no other demographical variables to be predictive of this disabling disease trend and it remains speculative whether baseline MRI activity or burden of disease will have prognostic power. The predictive value of our categorical classification remains to be validated with further follow up of our patients. Due to the nature of relapsing-remitting multiple sclerosis, it is likely that at least some misclassifications will have occurred, as shown in previous attempts on disease course assignments.<sup>26</sup> However, compared with confirmed progression outcomes, disease trend analyses maximise serial data utilisation and eliminate a substantial proportion of cases with erroneous progression.

### Conclusions

The two relatively new outcome methods included in our study are the summary measure statistic (AUC) and the categorical classification of disease trends on the EDSS. There are several prerequisites for these techniques. Firstly, acquisition of study data needs to be adequate, as the methods rely on most of the serial data points being available from each patient and their accuracy depends at least partly on sampling frequency. Secondly, these are not new scales, but merely analytical tools that inherit any drawbacks particular to the scale employed. However, the techniques described may prove even more powerful with the development of new, more reliable, and responsive rating measures (for example, multiple sclerosis functional composite; UK neurological disability scale).<sup>27-29</sup> The optimum sampling frequency for these rating scales needs to be determined, as the pay off between accuracy and noise is not yet known.<sup>30</sup>

In conclusion, we have found in our cohort of patients treated with placebo that the EDSS change over 2 years was small, although improved data utilisation could be achieved by performing summary measure AUC analysis, which is clinically more meaningful in relapsing-remitting multiple sclerosis. In addition, the employment of confirmed progression end points can be erroneous and misleading in this population due to a substantial proportion (about 50%) of patients who recover after prolonged periods of deteriorating disability. Thus, the interpretations of putative treatment benefits on disease progression based on these outcome measures must be reconsidered. An alternative approach, using all the serial EDSS

data to plot and categorise the in-trial disability course for each individual patient, was carried out. This disclosed a marked heterogeneity in which only a small proportion of patients showed sustained in-trial worsening or improvement. We found no useful clinical correlates which would enable prediction of these disability courses and the long term accuracy of this method remains to be determined.

Despite the advent of therapeutic trials in progressive multiple sclerosis,<sup>31</sup> the treatment of relapsing-remitting patients remains important, particularly in view of recent pathological<sup>32,33</sup> and MRI<sup>34,35</sup> evidence of axonal damage and atrophy early in the disease. Early clinical observations<sup>36,37</sup> and more recent MRI<sup>38</sup> and epidemiological data<sup>20</sup> testify to the importance of disease activity levels in the first few years of the disease for predicting the risks of significant disability in later years. Although determining outcome by exacerbation count is commonly used, it is relatively crude<sup>6</sup> and confounded by difficulties in relapse definitions<sup>8</sup> as well as quantification.<sup>7</sup> Although there is no substitute for prolonged trials to ascertain long term therapeutic benefits, in practice, studies led by the pharmaceutical industry are unlikely to be lengthy. Moreover, despite its lack of responsiveness<sup>39</sup> and the variable contributions of the different functional system scores towards disease progression,<sup>12</sup> the use of EDSS will probably remain prevalent. Therefore, improvement in data analysis methodology is also essential. Finally, the behaviour of patients treated with placebo in treatment trials can have a profound effect on the analysis of trial results and should be carefully examined before concluding that apparent treatment benefits are solely due to improvements in the actively treated trial arms.<sup>11</sup>

Thanks are due to the investigators of the Copolymer 1 Multiple Sclerosis Study Group and the PRISMS Study Group who performed the neurological assessments and who, with the permission of Teva Pharmaceutical Industries and Ares Serono International SA, agreed to supply us with the clinical data. The full lists of the trial investigators can be found in references 2 and 5. The authors were coinvestigators of both the PRISMS study, and the multinational, multicentre, randomised, double blind placebo controlled study, extended by open label treatment, to study the effect of glatiramer acetate (Copaxone) on disease activity as measured by cerebral MRI in patients with relapsing-remitting multiple sclerosis.

- 1 The IFNB Multiple Sclerosis Study Group. Interferon  $\beta$ -1b is effective in relapsing-remitting multiple sclerosis. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* 1993;43:655-61.
- 2 Johnson KP, Brooks BR, Cohen JA, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind, placebo-controlled trial. *Neurology* 1995;45:1268-76.
- 3 Jacobs LD, Cookfair DL, Rudick RA, et al. Intramuscular interferon  $\beta$ -1a for disease progression in relapsing multiple sclerosis. *Ann Neurol* 1996;39:285-94.
- 4 Fazekas F, Deisenhammer F, Strasser-Fuchs S, et al. Randomised placebo-controlled trial of monthly intravenous immunoglobulin therapy in relapsing-remitting multiple sclerosis. *Lancet* 1997;349:589-93.
- 5 PRISMS Study Group. Randomised double-blind placebo-controlled study of interferon  $\beta$ -1a in relapsing/remitting multiple sclerosis. *Lancet* 1998;352:1498-504.
- 6 Whitaker JN, McFarland HF, Rudge P, et al. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Multiple Sclerosis* 1995;1:37-47.
- 7 Hughes RAC, Sharrack B. More immunotherapy for multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1996;61:239-41.
- 8 Liu C, Blumhardt LD. Assessing relapses in treatment trials of relapsing and remitting multiple sclerosis: can we do better? *Multiple Sclerosis* 1999;5:22-8.

- 9 Weinschenker BG, Bass B, Rice GPA, *et al.* The natural history of multiple sclerosis: a geographically based study. 1. Clinical course and disability. *Brain* 1989;112:133–46.
- 10 Runmarker B, Andersen O. Prognostic factors in a multiple sclerosis incidence cohort with 25 years of follow-up. *Brain* 1993;116:117–34.
- 11 Noseworthy JH. Are placebo-controlled clinical trials still ethical in multiple sclerosis? In: Thompson AJ, Polman C, Hohlfeld R, eds. *Multiple sclerosis: clinical challenges and controversies*. London: Martin Dunitz, 1997:177–93.
- 12 Rice G, Ebers G. Interferons in the treatment of multiple sclerosis. Do they prevent the progression of the disease? *Arch Neurol* 1998;55:1578–80.
- 13 The IFNB Multiple Sclerosis Study Group and the University of British Columbia MS/MRI Analysis Group. Interferon  $\beta$ -1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology* 1995;45:1277–85.
- 14 Rudick RA, Goodkin DE, Jacobs LD, *et al.* Impact of interferon  $\beta$ -1a on neurologic disability in relapsing multiple sclerosis. *Neurology* 1997;49:358–63.
- 15 Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33:1444–52.
- 16 Liu C, Li Wan Po A, Blumhardt LD. Summary measure statistic for assessing the outcome of treatment trials in relapsing-remitting multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1998;64:726–9.
- 17 Weinschenker BG, Rice GPA, Noseworthy JH, *et al.* The natural history of multiple sclerosis: a geographically based study. 4. Applications to planning and interpretation of clinical therapeutic trials. *Brain* 1991;114:1057–67.
- 18 Ellison GW, Myers LW, Leake BD, *et al.* Design strategies in multiple sclerosis clinical trials. *Ann Neurol* 1994;36:S108–12.
- 19 Wingerchuk DM, Noseworthy JH, Weinschenker BG. Clinical outcome measures and rating scales in multiple sclerosis trials. *Mayo Clin Proc* 1997;72:1070–9.
- 20 Weinschenker BG, Bass B, Rice GPA, *et al.* The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. *Brain* 1989;112:1419–28.
- 21 Liu C, Blumhardt LD, Li Wan Po A. Reply to correspondence by Simonian NA, McAllister A, Hull J. Summary measure statistic for assessing the outcome of treatment trials in relapsing-remitting multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1999;66:553–4.
- 22 Weinschenker BG, Issa M, Baskerville J. Meta-analysis of the placebo-treated groups in clinical trials of progressive MS. *Neurology* 1996;46:1613–19.
- 23 Goodkin DE. Interferon  $\beta$  therapy for multiple sclerosis. *Lancet* 1998;352:1486–7.
- 24 McAlpine D, Compston N. Some aspects of the natural history of disseminated sclerosis. *Q J Med* 1952;82:135–67.
- 25 Weinschenker BG, Issa M, Baskerville J. Long-term and short-term outcome of multiple sclerosis. A 3 year follow up study. *Arch Neurol* 1996;53:353–8.
- 26 Goodkin DE, Hertsgaard D, Rudick RA. Exacerbation rates and adherence to disease type with a prospectively followed-up population with multiple sclerosis. Implications for clinical trials. *Arch Neurol* 1989;46:1107–12.
- 27 Rudick RA, Antel J, Confavreux C, *et al.* Recommendations from the clinical outcome assessment task force of the National Multiple Sclerosis Society. *Ann Neurol* 1997;42:379–82.
- 28 Cutter GR, Baier ML, Rudick RA, *et al.* Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* 1999;122:871–82.
- 29 Sharrack B, Hughes RAC. The Guy's neurological disability scale (GNDS): a new disability measure for multiple sclerosis. *Multiple Sclerosis* 1999;5:223–33.
- 30 Rudick RA, Antel J, Confavreux C, *et al.* Clinical outcomes assessment in multiple sclerosis. *Ann Neurol* 1996;40:469–79.
- 31 European Study Group on Interferon  $\beta$ -1b in Secondary Progressive MS. Placebo-controlled multicentre randomised trial of interferon  $\beta$ -1b in treatment of secondary progressive multiple sclerosis. *Lancet* 1998;352:1491–7.
- 32 Ferguson B, Matyszak MK, Esiri MM, *et al.* Axonal damage in acute multiple sclerosis lesions. *Brain* 1997;120:393–9.
- 33 Trapp BD, Peterson J, Ransohoff RM, *et al.* Axonal transection in the lesions of multiple sclerosis. *N Engl J Med* 1998;338:278–85.
- 34 Fu L, Matthews PM, De Stefano N, *et al.* Imaging axonal change of normal-appearing white matter in multiple sclerosis. *Brain* 1998;121:103–13.
- 35 Liu C, Edwards S, Gong Q, *et al.* Three dimensional MRI estimates of brain and spinal cord atrophy. *J Neurol Neurosurg Psychiatry* 1999;66:323–30.
- 36 Müller R. Studies on disseminated sclerosis: with special reference to symptomatology, course and progress. *Acta Med Scand* 1949;133(suppl 222):1–214.
- 37 Alexander L, Berkeley AW, Alexander AM. Prognosis and treatment of multiple sclerosis: quantitative nosometric study. *JAMA* 1958;166:1943–9.
- 38 Morrisey SP, Miller DH, Kendall BE, *et al.* The significance of brain magnetic resonance imaging abnormalities at presentation with clinically isolated syndromes suggestive of multiple sclerosis. A 5 year follow up study. *Brain* 1993;116:135–46.
- 39 Sharrack B, Hughes RAC, Soudain S, *et al.* The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain* 1999;122:141–59.