

# Randomised, double blind, placebo controlled study of interferon $\beta$ -1a in relapsing-remitting multiple sclerosis analysed by area under disability/time curves

Clarence Liu, Lance D Blumhardt

## Abstract

**Objectives**—The commonly employed outcome measures on disability and relapse rates in treatment trials of relapsing-remitting multiple sclerosis have well demonstrated sensitivity to treatment effects, but their clinical interpretation is problematic. An alternative method of analysis, which is more clinically meaningful and statistically appropriate to a condition with a fluctuating disease course, uses the summary measure statistic “area under the disability/time curve (AUC)”, to estimate each patient’s total in trial morbidity experience.

**Methods**—The AUC technique was applied in an intention to treat analysis of serial disability data derived from the expanded disability status scale (EDSS), the Scripps neurologic rating scale (SNRS), and the ambulation index (AI), collected during a double blind, randomised, placebo controlled, phase III trial of subcutaneous interferon  $\beta$ -1a (INF $\beta$ -1a) in relapsing-remitting multiple sclerosis (PRISMS Study). The results were compared with the often quoted “conventional” end point of mean change in rating scores from baseline to trial completion. Analyses were also carried out on subgroups with entry EDSS stratified above and below 3.5.

**Results**—EDSS data analysed by AUC normalised to baseline scores disclosed that both doses of INF $\beta$ -1a (22 or 44  $\mu$ g) were superior to placebo ( $p=0.008$  and  $0.013$ , respectively). In addition, the high dose (44  $\mu$ g) was more beneficial than placebo using SNRS ( $p=0.038$ ) and AI data ( $p=0.039$ ). AUC analysis of SNRS scores also showed that for patients with baseline EDSS $>3.5$ , the 44  $\mu$ g (but not the 22  $\mu$ g) dose was more advantageous than placebo ( $p=0.028$ ).

**Conclusions**—Summary measure analysis using the AUC of serial disability/time plots, confirms and extends the results of conventional end point analysis of disability from the PRISMS Study data. AUC evaluations show that high dose INF $\beta$ -1a (44  $\mu$ g three times weekly) was beneficial on all of the clinical rating scale scores used in this study. This method provides a statistically powerful and clinically meaningful assessment of treatment effects on

**in trial disability in patients with multiple sclerosis with fluctuating and highly heterogeneous disease courses.**

(J Neurol Neurosurg Psychiatry 1999;67:451-456)

**Keywords:** multiple sclerosis; outcome measures; interferon  $\beta$ -1a

In the past 5 years, immunomodulatory therapies for relapsing-remitting multiple sclerosis have achieved a useful reduction in relapse rate, but less clear cut benefits on permanent disability, based on the outcome measures employed.<sup>1-5</sup> These measures have been associated either with a change in rating scores from study baseline to completion, or with “confirmed progression” as defined by a certain increase in disability scores (for example, the expanded disability status scale, EDSS)<sup>6</sup> at two visits 3 or 6 months apart. Interpretation of these measures is difficult, as the first end point ignores the instability and variance associated with two snapshot assessments in time, as well as the fluctuating disabilities that commonly occur in relapsing-remitting multiple sclerosis, whereas the so called confirmed progression (and its graphical depiction using Kaplan-Meier survival curves) includes an unknown number of erroneous treatment failures (that is, cases with recovery to baseline after satisfying so called progression criteria). End points which are both clinically and statistically meaningful and which incorporate the amount of disability associated with each attack into the overall disability calculations<sup>7</sup> would be preferred.

One method of analysis that indexes all in trial morbidity changes is the area under the disability/time curve (AUC).<sup>8</sup> This technique is appropriate for the mainly transient disability experienced early in the disease course in relapsing-remitting multiple sclerosis and should be more responsive to change as it utilises all the collected serial data. In the present study, we have reanalysed with AUCs the clinical rating scores obtained in a randomised, double blind, placebo controlled trial of interferon  $\beta$ -1a (INF $\beta$ -1a: Rebif<sup>®</sup>, Ares Serono) given subcutaneously in relapsing-remitting multiple sclerosis, recently published by the PRISMS (prevention of relapses and disability by interferon-beta 1a subcutaneously in multiple sclerosis) Study Group.<sup>9</sup> The results are compared with a conventional method of disability analysis (2 year change in clinical rat-

Division of Clinical Neurology, Department of Medicine, University Hospital, Queen’s Medical Centre, Nottingham NG7 2UH, 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 1 February 1999 and in revised form 19 April 1999  
Accepted 21 May 1999

**Table 1** Total cohort (n=560) disability changes as measured by (A) EDSS, (B) SNRS, and (C) AI, analysed using 2 year disability difference and AUC<sup>CHANGE</sup> for all visits: median (interquartile difference) and mean (SD)

	Two year difference		AUC <sup>CHANGE</sup> (in EDSS-years)	
	Median	Mean	Median	Mean
<b>(A) EDSS:</b>				
IFNβ-1a 44 µg (3×/week)	0 (1.5)	+0.24 (1.13)	+0.06 (1.67)	+0.33 (1.63)
IFNβ-1a 22 µg (3×/week)	0 (1.5)	+0.23 (1.29)	+0.05 (1.76)	+0.36 (1.89)
Placebo	+0.5 (1.0)	+0.48 (1.27)	+0.48 (1.75)	+0.81 (1.95)
	Two year difference		AUC <sup>CHANGE</sup> (in SNRS-years)	
	Median	Mean	Median	Mean
<b>(B) SNRS:</b>				
IFNβ-1a 44 µg (3×/week)	-1 (10)	-1.10 (8.02)	+0.17 (13.12)	-0.84 (11.65)
IFNβ-1a 22 µg (3×/week)	0 (9.5)	-1.74 (9.13)	-0.25 (9.63)	-1.46 (12.21)
Placebo	-2 (7.5)	-3.13 (9.10)	-1.68 (11.83)	-3.29 (12.39)
	Two year difference		AUC <sup>CHANGE</sup> (in AI-years)	
	Median	Mean	Median	Mean
<b>(C) AI:</b>				
IFNβ-1a 44 µg (3×/week)	0 (1)	+0.24 (0.96)	0 (1.12)	+0.25 (1.31)
IFNβ-1a 22 µg (3×/week)	0 (1)	+0.46 (1.25)	+0.17 (0.97)	+0.53 (1.60)
Placebo	0 (1)	+0.44 (1.30)	+0.11 (3.05)	+0.65 (2.51)

ing scores from baseline to end of study). We did not include the confirmed progression end point as it is not strictly comparable with the AUC, because it only utilises a minority of the available disability data in the proportion of patients who reach the end point concerned. In addition, AUC provides no information about the direction of disability change. Thus confirmed progression will be evaluated by comparison with trends established from individual disability/time plots in a separate study.

## Methods

### PATIENTS

A total of 560 patients with relapsing-remitting multiple sclerosis were enrolled and randomised to 44 µg (12 MIU) or 22 µg (6 MIU) IFNβ-1a or placebo, administered three times weekly by subcutaneous injection (totalling 132, 66, and 0 µg IFNβ-1a per week, respectively). Details of the patient cohort, exclusion criteria, and clinical results have been reported in an earlier publication by the PRISMS Study Group.<sup>9</sup>

Patients were assessed neurologically and scored on the EDSS, Scripps neurologic rating scale (SNRS)<sup>10</sup> and the ambulation index (AI)<sup>11</sup> at the start of the trial and at each visit, by an examining neurologist blinded to the treatment category. In 14 centres, 316 patients had scheduled assessments carried out at 3 monthly appointments. In another six centres, additional monthly scheduled visits were performed for the first 9 months in 205 patients. In one further centre, 39 patients had scheduled monthly assessments throughout the study. Extra evaluations (n=592) were carried out at unscheduled patient initiated visits associated with relapses. Total disability data sets were achieved in 533 patients (95%).

### ANALYSIS

Conventional analysis (2 year disability difference) was carried out by comparing treatment effects on the changes in EDSS, SNRS, and AI, between trial entry and completion. AUCs

were calculated by two methods to obtain the AUC<sup>SUM</sup> and the AUC<sup>CHANGE</sup>. For AUC<sup>SUM</sup>, the total area under the disability/time curve throughout the trial was determined.<sup>8,12</sup> For AUC<sup>CHANGE</sup>, the AUC<sup>SUM</sup> was normalised to the baseline score by subtracting the area defined by the product of the initial rating score and the study period (see appendix). Between EDSS of 5.5 to 7.0, each 0.5 point increment was rescaled to 1.0 point to adjust for non-linearity of the EDSS.<sup>13,14</sup>

Three types of data analyses were performed. In the initial analyses (combined data), all 7060 datapoints from the scheduled and unscheduled visits of 560 patients were included in the AUC calculations (intention to treat analysis). The trapezium rule for determining the AUC was applied throughout,<sup>12</sup> as most objectively confirmed attacks had only one additional neurological assessment and the speed and timing of relapse onset<sup>15</sup> and offset<sup>16</sup> would be difficult to define with certainty if other techniques were employed. Secondly, separate analyses (scheduled visit data) were carried out using solely the 6468 datapoints obtained at routine appointments. Thirdly, due to the different estimated disease course characteristics and DSS score staying times of patients with DSS≤3 and DSS≥4 in natural history series,<sup>15</sup> analyses of subjects classified by baseline EDSS≤3.5 (n=466) or EDSS>3.5 (n=94) were attained.

### STATISTICS

For both conventional and AUC data, an analysis of variance (ANOVA) model was employed with factors for treatment and centre. One degree of freedom contrasts from the ANOVA model were used to compare the treatment groups in a pairwise fashion. As there was a strong effect of baseline entry scores, all further analyses of AUC<sup>SUM</sup> were carried out with baseline disability (EDSS, SNRS, or AI) as a covariant.<sup>13</sup> ANOVAs on the ranks (Kruskal-Wallis tests) were performed to determine consistency of results and validity of the parametric (ANOVA) conclusions. As our data were not normally distributed (Shapiro-Wilk *W* test), *p* values on treatment comparisons were obtained from ANOVAs on the ranks. Mean estimates with 95% confidence intervals (95% CIs were also calculated to compare treatment effects with placebo.

## Results

### TOTAL COHORT ANALYSES WITH COMBINED DATA EDSS

"2 year EDSS difference" demonstrated significant benefits favouring the 22 µg dose (p=0.026) and a tendency in favour of 44 µg (p=0.052) over placebo. AUC<sup>SUM</sup> calculations similarly showed that 22 µg IFNβ-1a was beneficial compared with placebo (p=0.046). Although the size of the treatment effect was similar, significance was not reached with the 44 µg dose (p=0.064).

The median AUC<sup>CHANGE</sup> was +0.06, +0.05, and +0.48 EDSS-year for 44 µg, 22 µg IFNβ-1a, and placebo, respectively (table 1 A). Both doses conferred significant advantages

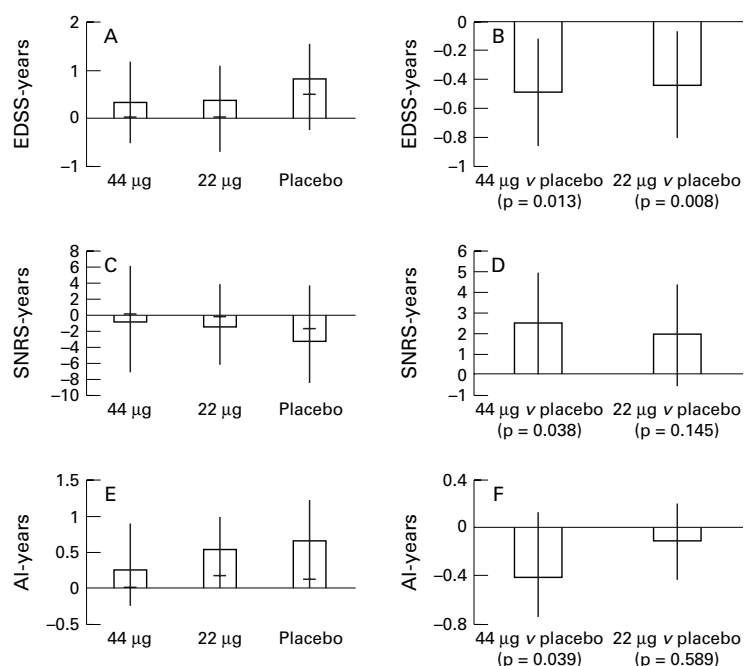


Figure 1  $AUC^{CHANGE}$  data obtained on total visits with intention to treat analysis for the entire cohort ( $n=560$ ).  $AUC^{CHANGE}$  EDSS: (A) basic statistics and (B) treatment comparisons in EDSS-years.  $AUC^{CHANGE}$  SNRS: (C) basic statistics and (D) treatment comparisons in SNRS-years.  $AUC^{CHANGE}$  AI: (E) basic statistics and (F) treatment comparisons in AI-years. (Basic statistics: histogram=group mean, horizontal bar=median, vertical line=interquartile range. Treatment comparisons: histogram=estimated mean, vertical line=95% CI,  $p$  value from ANOVA on ranks).

over placebo ( $p=0.013$  and  $0.008$  for  $44 \mu\text{g}$  and  $22 \mu\text{g}$  IFN $\beta$ -1a, respectively), with an estimated treatment effect of  $-0.5$  EDSS-year for each dose (95% CI  $-0.9$  to  $-0.1$  and  $-0.8$  to  $-0.1$ , respectively) (fig 1 A-B).

SNRS

There were no significant group differences in efficacy based on either “2 year SNRS difference”, or  $AUC^{SUM}$  analysis. With  $AUC^{CHANGE}$ , the  $44 \mu\text{g}$  treatment arm improved by a median of  $+0.17$  SNRS-year compared to deteriorations of  $-0.25$  and  $-1.68$  SNRS-year for the 22

$\mu\text{g}$  dose and placebo cohorts, respectively (table 1 B). There were significant effects in favour of  $44 \mu\text{g}$  IFN $\beta$ -1a over placebo ( $p=0.038$ ), with an estimated mean benefit of  $+2.5$  SNRS-year (95% CI  $+0.1$  to  $+4.9$ ) (fig 1 C-D).

AI

There were no significant group differences for either the “2 year AI difference”, or  $AUC^{SUM}$ . On  $AUC^{CHANGE}$ , the  $44 \mu\text{g}$  dose (median of 0 AI-year) was significantly superior to placebo (median worsening of  $+0.11$  AI-year) ( $p=0.039$ ) with a mean benefit of  $-0.4$  AI-year (95% CI  $-0.7$  to  $-0.1$ ) (table 1 C, fig 1 F).

TOTAL COHORT ANALYSES WITH SCHEDULED VISIT DATA

Utilising data from scheduled visits only,  $AUC^{SUM}$  analysis on EDSS scores disclosed a trend favouring  $22 \mu\text{g}$  IFN $\beta$ -1a over placebo ( $p=0.058$ ).  $AUC^{CHANGE}$  calculations showed that on EDSS scores, there were significant benefits of both doses over placebo ( $p=0.024$  and  $0.014$  for  $44 \mu\text{g}$  and  $22 \mu\text{g}$ , respectively). For SNRS and AI data, treatment with IFN $\beta$ -1a  $44 \mu\text{g}$  was also superior to placebo.

SUBGROUP ANALYSES

EDSS

For patients with entry EDSS  $\leq 3.5$ , treatment comparisons showed  $22 \mu\text{g}$  to be better than placebo, for both “2 year EDSS difference” ( $p=0.016$ ) and  $AUC^{SUM}$  analyses ( $p=0.043$ ). Using  $AUC^{CHANGE}$ , there were significant effects in favour of both treatment doses compared with placebo ( $p=0.036$  and  $0.016$  for  $44 \mu\text{g}$  and  $22 \mu\text{g}$  IFN $\beta$ -1a), with estimated benefits of  $-0.4$  and  $-0.5$  EDSS-year, respectively (95% CI  $-0.8$  to  $0$ ;  $-0.8$  to  $-0.1$ ) (table 2 A, fig 2 A-B).

For patients with baseline EDSS  $> 3.5$ , neither conventional “2 year EDSS difference”, nor AUC analyses, showed significant differences between treatment arms, although

Table 2 Low (EDSS  $\leq 3.5$ ,  $n=466$ ) and high (EDSS  $> 3.5$ ,  $n=94$ ) entry disability subgroups: Disability changes as measured by (A) EDSS, (B) SNRS and (C) AI, analysed using 2 year disability difference and  $AUC^{CHANGE}$  for all visits: median (interquartile difference) and mean (SD)

	Entry EDSS $\leq 3.5$ subgroup				Entry EDSS $> 3.5$ subgroup			
	Two year difference		$AUC^{CHANGE}$ (in EDSS-years)		Two year difference		$AUC^{CHANGE}$ (in EDSS-years)	
	Median	Mean	Median	Mean	Median	Mean	median	mean
<b>(A) EDSS:</b>								
IFN $\beta$ -1a $44 \mu\text{g}$ (3 $\times$ /week)	0 (1.0)	+0.21 (1.13)	+0.04 (1.79)	+0.23 (1.58)	0 (1.0)	+0.38 (1.15)	+0.26 (1.21)	+0.83 (1.79)
IFN $\beta$ -1a $22 \mu\text{g}$ (3 $\times$ /week)	0 (1.0)	+0.12 (1.27)	+0.05 (1.79)	+0.20 (1.79)	+0.5 (1.0)	+0.69 (1.30)	0 (2.10)	+1.07 (2.20)
Placebo	+0.5 (1.25)	+0.37 (1.18)	+0.38 (1.66)	+0.64 (1.84)	+0.5 (2.0)	+1.14 (1.59)	+1.07 (3.00)	+1.78 (2.28)
Two year difference		$AUC^{CHANGE}$ (in SNRS-years)		Two year difference		$AUC^{CHANGE}$ (in SNRS-years)		
Median	Mean	Median	Mean	Median	Mean	median	mean	
<b>(B) SNRS:</b>								
IFN $\beta$ -1a $44 \mu\text{g}$ (3 $\times$ /week)	0 (7)	-1.34 (7.66)	0 (12.68)	-1.11 (11.54)	-3 (16)	+0.07 (9.67)	+0.81 (18.74)	+0.46 (12.25)
IFN $\beta$ -1a $22 \mu\text{g}$ (3 $\times$ /week)	0 (9)	-1.57 (9.15)	-0.32 (9.51)	-1.15 (11.49)	-2 (14)	-2.49 (9.12)	-0.05 (18.93)	-2.82 (15.09)
Placebo	-1 (7)	-2.49 (8.48)	-1.46 (9.86)	-2.34 (11.22)	-4 (12)	-7.00 (11.68)	-7.86 (19.65)	-8.67 (16.89)
Two year difference		$AUC^{CHANGE}$ (in AI-years)		Two year difference		$AUC^{CHANGE}$ (in AI-years)		
Median	Mean	Median	Mean	Median	Mean	median	mean	
<b>(C) AI</b>								
IFN $\beta$ -1a $44 \mu\text{g}$ (3 $\times$ /week)	0 (1)	+0.19 (0.93)	0 (0.86)	+0.16 (1.25)	0 (1)	+0.5 (1.07)	+0.18 (1.61)	+0.69 (1.54)
IFN $\beta$ -1a $22 \mu\text{g}$ (3 $\times$ /week)	0 (1)	+0.34 (1.13)	+0.10 (0.85)	+0.41 (1.38)	+1 (1)	+0.97 (1.59)	+0.61 (1.57)	+1.10 (2.27)
Placebo	0 (1)	+0.29 (1.03)	0 (1.27)	+0.48 (1.44)	+1 (1)	+1.36 (2.18)	+1.13 (1.81)	+1.65 (2.34)

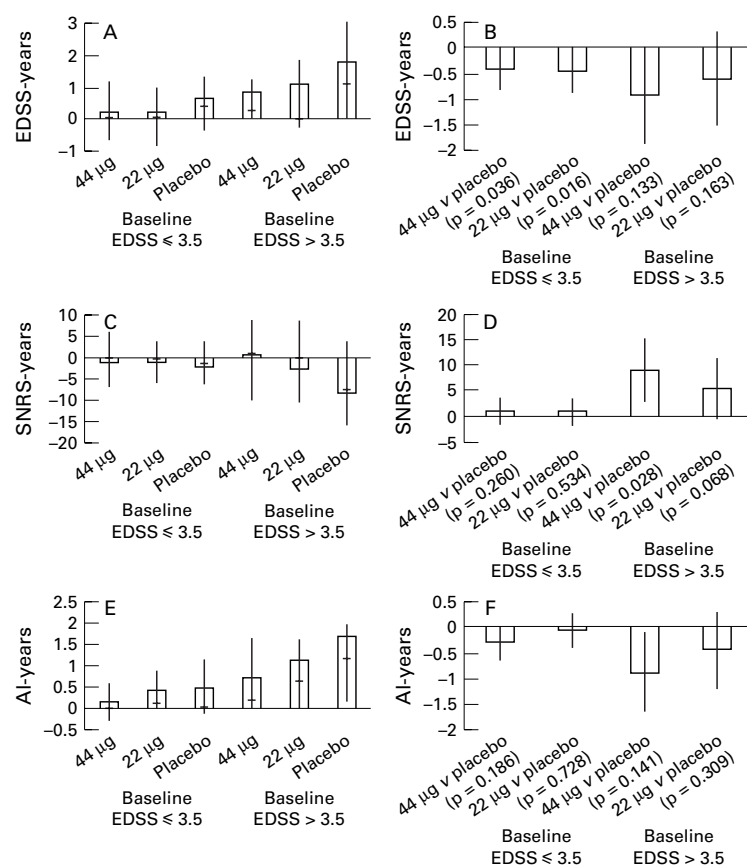


Figure 2  $AUC^{CHANGE}$  data obtained on total visits with intention to treat analysis for the subgroups of low ( $EDSS \leq 3.5$ ,  $n=466$ ) and high ( $EDSS > 3.5$ ,  $n=94$ ) entry disability.  $AUC^{CHANGE}$  EDSS (A) basic statistics and (B) treatment comparisons in EDSS-years.  $AUC^{CHANGE}$  SNRS (C) basic statistics and (D) treatment comparisons in SNRS-years.  $AUC^{CHANGE}$  AI (E) basic statistics and (F) treatment comparisons in AI-years. (Basic statistics: histogram=group mean, horizontal bar=median, vertical line=interquartile range. Treatment comparisons: histogram=estimated mean, vertical line=95% CI, p value from ANOVA on ranks).

$AUC^{CHANGE}$  showed an estimated mean benefit for the 44 μg dose over placebo of -1.0 EDSS-year (95% CI -1.9 to 0) (fig 2 A-B).

### SNRS

For patients with entry  $EDSS \leq 3.5$ , neither “2 year SNRS difference”,  $AUC^{SUM}$ , nor  $AUC^{CHANGE}$  showed any significant distinctions between treatment arms (table 2 B, fig 2 C-D).

For patients with a baseline  $EDSS > 3.5$ , there was a significant effect in favour of the 44 μg dose over placebo using conventional “2 year SNRS difference” ( $p=0.016$ ).  $AUC^{SUM}$  did not show differences between treatment arms.  $AUC^{CHANGE}$  confirmed significant effects of 44 μg IFNβ-1a versus placebo ( $p=0.028$ ) with a mean benefit of +9.0 SNRS-years (95% CI +2.9 to +15.1) (fig 2 C-D).

### AI

Neither “2 year AI difference”,  $AUC^{SUM}$ , nor  $AUC^{CHANGE}$  showed any significant treatment effects over placebo for either subgroup (table 2 C, fig 2 E-F). However, for  $AUC^{CHANGE}$ , the 44 μg dose conferred an estimated mean saving of -0.9 AI-year over placebo (95% CI -1.6 to -0.1).

### Subgroup scheduled visit data analyses

Treatment comparisons between subgroups using data restricted to scheduled visits showed significant benefits of both treatment doses for patients with entry  $EDSS \leq 3.5$  on EDSS assessments ( $p=0.021$  to  $0.043$ ), and favourable effects on SNRS scores with high dose 44 μg IFNβ-1a over placebo in patients with  $EDSS > 3.5$  ( $p=0.034$ ).

### Discussion

Whereas it is generally agreed that the primary outcome in phase III immunomodulatory trials in relapsing-remitting multiple sclerosis must be clinical,<sup>17</sup> commonly employed end points associated with both disability and relapse assessment can be difficult to define and interpret.<sup>8, 16</sup> Meaningful clinical milestones, such as the conversion to secondary progression,<sup>18, 19</sup> are not realistic end points for relapsing-remitting multiple sclerosis cohorts within relatively short study periods of 2 to 3 years. Currently available clinical rating scales are variously flawed,<sup>20-22</sup> with the EDSS being the most criticised.<sup>13, 14, 20, 23, 24</sup> Other scales (including SNRS and AI) have been less thoroughly evaluated.<sup>25</sup> We have previously proposed that the AUC summary measure statistic may solve some of these problems in a clinically meaningful way by indexing and accounting for the total morbidity change experienced by patients during the course of a trial.<sup>8</sup>

In this paper, we have analysed data from the PRISMS trial using two AUC methods,  $AUC^{SUM}$  and  $AUC^{CHANGE}$ .  $AUC^{SUM}$  summates the total disability data serially over the study period and data independence is assured (the independence is lost in analyses involving any change in disability scores as each data point is influenced by its preceding values). However, it would be expected to be insensitive to small changes in short trials if there were a large disability range in the cohort at baseline. On the other hand,  $AUC^{CHANGE}$  obtained by normalising the  $AUC^{SUM}$  value to the baseline rating is sensitive to in trial changes and has been previously utilised in a neurorehabilitation pilot study.<sup>26</sup> However, neurological improvements and deteriorations may cancel and the technique is susceptible to unstable scores at trial entry, a problem partly resolved by ensuring a stable run in period.

In the present study, comparison of the two AUC techniques showed that  $AUC^{SUM}$  essentially confirmed the “2 year disability difference” analysis (treatment benefit over placebo for EDSS, but not for SNRS or AI data) without improved responsiveness. This can be attributed to the small in study changes relative to the wide range of disability ( $EDSS$  0-5) at baseline. By contrast, the increased sensitivity of  $AUC^{CHANGE}$  for detecting positive therapeutic effects not only showed savings in terms of EDSS-years for both IFNβ-1a doses compared with placebo, but also significant beneficial effects in favour of the high dose (44 μg three times weekly), using SNRS and AI data.

AUC analysis has several advantages over conventional techniques for this type of study. The method can be applied to any clinical rat-



ing scale employed. Because it incorporates all available serial data, it should be statistically more powerful than end points of conventional mean disability change. By accounting for both the magnitude and duration of relapses as well as improvements and progressive deteriorations, it is more clinically meaningful, as it provides a measure of the patient's total disability experience. Furthermore, cost effectiveness can be assessed in terms of disability-years. These factors allow a more complete intention to treat evaluation.

We examined the neurological scores from all visits (combined data), as well as those assessed only from scheduled appointments (scheduled visit data). The second was performed to reduce any sampling bias due to the reduced mean relapse rate in the treatment arms compared with placebo, and to separate out, at least in part, the effects of transient disability changes. The results in favour of therapy were similar with or without the inclusion of data from unscheduled visits.

We analysed our higher disability cohort separately in this study because there is accumulating evidence that patients with active relapsing-remitting multiple sclerosis and an EDSS >3.5 (when gait dysfunction ensues) are at a higher risk for disability progression.<sup>27</sup> These patients have a tendency to deteriorate more rapidly<sup>13</sup> and to be unresponsive to steroids and standard doses of IFN $\beta$ .<sup>28</sup> The AUC<sup>CHANGE</sup> analysis of this subgroup disclosed that patients on placebo experienced strikingly more disability on EDSS, SNRS, and AI than their treated counterparts. AUC<sup>CHANGE</sup> confirmed the conventional end point of 2 year SNRS difference in showing significant treatment effects only for the 44  $\mu$ g dose. Furthermore, from the estimated means and confidence intervals, trends were apparent for similar dose effects with the EDSS and AI data, although statistical significance was not reached due to the relatively few subjects. Overall, these results support the notion that subjects with significant disability require higher doses than patients at the lower end of the EDSS (with mainly impairment), to obtain similar benefits.<sup>27</sup>

The differential outcomes of the two treatment doses on disability as assessed by the different rating scales are intriguing and difficult to explain. They may be confounded by the modest correlations within individual patients across changes in rating scores,<sup>29</sup> and the relatively lack of comparative data on SNRS and AI.<sup>25</sup> Moreover, in the PRISMS Study cohort, the SNRS scoring system generates a more continuous dataset than the EDSS or AI and thus, for the relatively small subgroup of subjects with baseline EDSS >3.5 (n=94), may be more sensitive to the treatment effects of the 44  $\mu$ g dose.

AUC analyses might be regarded as having certain disadvantages. For example, the method summarises all the disability experienced by a patient during a trial, but does not distinguish between permanent neurological dysfunction and the transient disability associated with relapses. However, we regard this as

being useful for assessing the effects of potential therapies in patients with relapsing-remitting multiple sclerosis in whom significant transient morbidity may occur over short periods despite the slow accumulation of irreversible disability.<sup>19,30</sup> AUC analyses also provide no information on disability trends over time: a subject with a period of improvement followed by deterioration may have the same AUC score as one with the opposite temporal sequence, or another in which the morbidity remains unaltered throughout. This issue, as well as the problem of erroneous treatment failures wrongly assigned by confirmed progression and Kaplan-Meier analyses, can only be solved by evaluating individual disability/time plots. The question of an optimal sampling rate, including concerns of the signal to noise of rating scales employed at high assessment frequency,<sup>17</sup> also needs to be considered. Although considered appropriate for analysing graded data,<sup>12</sup> AUC analyses do not eliminate problems peculiar to the clinical rating tools utilised, as discussed previously. However, these difficulties likewise apply to the disability end points presently utilised in clinical trials. It has been suggested that improved clinical rating tools, preferably with continuous scales, are required (for example, maximum ambulatory distance, timed 8 metre walk, nine hole peg test, paced auditory serial addition task, questionnaire based assessments).<sup>31-35</sup> The numerical data derived from such scales is likely to be even more informative when subjected to AUC analyses.

CL and LDB were investigators in the PRISMS study. Thanks are due to Florilene Dupont and Melvin Olsen of Ares Serono for help in data analysis.

## Appendix

**AUC<sup>SUM</sup>**: This is a summation of the total areas (including baseline) under the curve between each pair of consecutive scores given by the trapezium rule. If disability scores ( $y_0, y_1, y_2, \dots, y_n$ ) are plotted versus their times of assessment ( $t_0, t_1, t_2, \dots, t_n$ ) totalling n+1 measurements  $y_i$  at times  $t_i$  ( $i=0, \dots, n$ ), then

$$AUC^{SUM} = 1/2 \sum_{i=0}^{n-1} (t_{i+1} - t_i)(y_i + y_{i+1})$$

**AUC<sup>CHANGE</sup>**: This is the difference between the areas summated by AUC<sup>SUM</sup> and the product of the baseline disability score  $y_0$  and the total time of study  $t_n$ , hence

$$AUC^{CHANGE} = AUC^{SUM} - (y_0)(t_n - t_0)$$

- 1 IFNB Multiple Sclerosis Study Group. Interferon  $\beta$ -1b is effective in relapsing-remitting multiple sclerosis. 1. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* 1993;43:655-61.
- 2 IFNB Multiple Sclerosis Study Group and the University of British Columbia multiple sclerosis/MRI Analysis Group. Interferon  $\beta$ -1b in the treatment of multiple sclerosis. Final outcome of the randomized controlled trial. *Neurology* 1995;45:1277-85.
- 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 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.
- 5 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.

- 6 Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;**33**:1444–52.
- 7 Hughes RAC, Sharrack B. More immunotherapy for multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1996;**61**:239–41.
- 8 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.
- 9 PRISMS (Prevention of relapses and disability by interferon- $\beta$  1a subcutaneously in multiple sclerosis) Study Group. Randomised, double-blind, placebo-controlled study of interferon- $\beta$  1a in relapsing-remitting multiple sclerosis: clinical results. *Lancet* 1998;**352**:1498–504.
- 10 Sipe JC, Knobler RL, Braheny SL, et al. A neurologic rating scale (NRS) for use in multiple sclerosis. *Neurology* 1984;**34**:1368–72.
- 11 Hauser SL, Dawson DM, Lehrich JR, et al. Intensive immunosuppression in progressive multiple sclerosis. A randomized, three-arm study of high-dose intravenous cyclophosphamide, plasma exchange, and ACTH. *N Engl J Med* 1993;**308**:173–80.
- 12 Matthews JNS, Altman DG, Campbell MJ, et al. Analysis of serial measurements in medical research. *BMJ* 1990;**300**:230–5.
- 13 Weinschenker BG, Bass B, Rice GPA, 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.
- 14 Wingerchuk DM, Noseworthy JH, Weinschenker BG. Clinical outcome measures and rating scales in multiple sclerosis trials. *Mayo Clin Proc* 1997;**72**:1070–9.
- 15 Alexander L, Berkeley AW, Alexander AM. Prognosis and treatment of multiple sclerosis: quantitative nosometric study. *JAMA* 1958;**166**:1943–9.
- 16 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.
- 17 Rudick R, Antel J, Confavreux C, et al. Clinical outcome assessment in multiple sclerosis. *Ann Neurol* 1996;**40**:469–79.
- 18 Confavreux C, Aimard G, Devic M. Course and prognosis of multiple sclerosis assessed by the computerized data processing of 349 patients. *Brain* 1980;**103**:281–300.
- 19 Runmarker B, Andersen O. Prognostic factors in a multiple sclerosis incidence cohort with 25 years of follow-up. *Brain* 1993;**116**:117–34.
- 20 Willoughby EW, Paty DW. Scales for rating impairment in multiple sclerosis: a critique. *Neurology* 1988;**38**:1793–8.
- 21 Whitaker JN, McFarland HF, Rudge P, et al. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Multiple Sclerosis* 1995;**1**:37–47.
- 22 Sharrack B, Hughes RAC. Clinical scales for multiple sclerosis. *J Neurol Sci* 1996;**135**:1–9.
- 23 Francis DA, Bain P, Swan AV, et al. An assessment of disability rating scales used in multiple sclerosis. *Arch Neurol* 1991;**48**:299–301.
- 24 Goodkin DE, Cookfair D, Wende K, et al. Inter- and intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke expanded disability status scale (EDSS). *Neurology* 1992;**42**:859–63.
- 25 Hobart JC. Measuring health outcomes in multiple sclerosis: why, which and how? In: Thompson AJ, Polman C, Hohlfeld R, eds. *Multiple sclerosis: clinical challenges and controversies*. London: Martin Dunitz, 1997:211–25.
- 26 Freeman J, Langdon DW, Hobart JC, et al. The long term effects of neurorehabilitation in multiple sclerosis: a longitudinal study. *J Neurol* 1997;**244**(suppl 3):S10.
- 27 Blumhardt LD, Paty DW, Hughes RAC, et al. Dosage effect of interferon  $\beta$ -1a (Rebif) in preventing relapses and progression of disability in relapsing-remitting multiple sclerosis with baseline EDSS>3.5. *J Neurol* 1998;**245**:371.
- 28 Weinstock-Guttman B, Kimbel RP, Cohen JA, et al. Treatment of transitional MS with cyclophosphamide and methylprednisolone followed by interferon  $\beta$ . *Neurology* 1997;**48**:A341.
- 29 Koziol JA, Frutos A, Sipe JC, et al. A comparison of two neurologic scoring instruments for multiple sclerosis. *J Neurol* 1996;**243**:209–13.
- 30 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.
- 31 Schwid SR, Goodman AD, Mattson DH, et al. The measurement of ambulatory impairment in multiple sclerosis. *Neurology* 1997;**49**:1419–24.
- 32 Goodkin DE, Hertsgaard D, Seminary J. Upper extremity function in multiple sclerosis: improving assessment sensitivity with box and block and nine hole peg tests. *Arch Phys Med Rehabil* 1998;**69**:850–4.
- 33 Gronwall D, Wrightson P. Delayed recovery of intellectual function after minor head injury. *Lancet* 1974;ii:605–9.
- 34 Rudick R, Antel J, Confavreux C, et al. Recommendations from the National Multiple Sclerosis Society clinical outcome assessment task force. *Ann Neurol* 1997;**42**:379–82.
- 35 Sharrack B, Hughes RAC, Soudain S. Guy's neurological disability scale. *J Neurol* 1996;**243**(suppl 2):S32.