Randomised, double blind, placebo controlled study of interferon β-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 β-1a (INFβ-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 INFβ-1a (22 or 44 µg) were superior to placebo (p= 0.008 and 0.013, respectively). In addition, the high dose (44 µg) 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 µg (but not the 22 µg) 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 INFβ-1a (44 µg 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.

Keywords: multiple sclerosis; outcome measures; interferon β-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. These measures have been associated with changes in milestone scores from baseline to completion, or with “confirmed progression” as defined by a certain increase in disability scores (for example, the expanded disability status scale, EDSS) 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 would be preferred.

One method of analysis that indexes all in trial morbidity changes is the area under the disability/time curve (AUC). 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 β-1a (IFNβ-1a: Rebif®, 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. The results are compared with a conventional method of disability analysis (2 year change in clinical rat-
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 AUCCHANGE for all visits: median (interquartile difference) and mean (SD)

<table>
<thead>
<tr>
<th>Method</th>
<th>Two year difference</th>
<th>( \text{AUC}_\text{CHANGE} ) (in EDSS-years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>(A) EDSS:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFN(^{-1})-la 44 µg (3/week)</td>
<td>0 (1.5)</td>
<td>+0.24 (1.13)</td>
</tr>
<tr>
<td>IFN(^{-1})-la 22 µg (3/week)</td>
<td>0 (1.5)</td>
<td>+0.23 (1.20)</td>
</tr>
<tr>
<td>Placebo</td>
<td>+0.5 (1.0)</td>
<td>+0.48 (1.27)</td>
</tr>
<tr>
<td>(B) SNRS:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFN(^{-1})-la 44 µg (3/week)</td>
<td>-1 (10)</td>
<td>-1.10 (8.02)</td>
</tr>
<tr>
<td>IFN(^{-1})-la 22 µg (3/week)</td>
<td>0 (9.5)</td>
<td>-1.74 (9.13)</td>
</tr>
<tr>
<td>Placebo</td>
<td>-2 (7.5)</td>
<td>-3.13 (9.10)</td>
</tr>
<tr>
<td>(C) AI:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFN(^{-1})-la 44 µg (3/week)</td>
<td>0 (1)</td>
<td>+0.24 (0.96)</td>
</tr>
<tr>
<td>IFN(^{-1})-la 22 µg (3/week)</td>
<td>0 (1)</td>
<td>+0.46 (1.25)</td>
</tr>
<tr>
<td>Placebo</td>
<td>0 (1)</td>
<td>+0.44 (1.30)</td>
</tr>
</tbody>
</table>

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 \( \text{AUC}_\text{SUM} \) were carried out with baseline disability (EDSS, SNRS, or AI) as a covariant. 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 ANALYSIS 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. \( \text{AUC}_\text{SUM} \) calculations similarly showed that 22 µg IFN\(^{-1}\)-la 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 \( \text{AUC}_\text{CHANGE} \) was +0.06, +0.05, and +0.48 EDSS-year for 44 µg, 22 µg IFN\(^{-1}\)-la, and placebo, respectively (table 1A). Both doses conferred significant advantages.
over placebo (p=0.013 and 0.008 for 44 µg and 22 µg IFNβ-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 AUCSUM analysis. With AUCCHANGE, the 44 µg 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 µg and placebo, for both “2 year SNRS difference” and AUC analyses, showed significant differences between treatment arms, although no AUC analyses, showed significant differences between treatment arms, although

**SUBGROUP ANALYSES**

**EDSS**

For patients with entry EDSS ≤3.5, treatment comparisons showed 22 µg to be better than placebo, for both “2 year EDSS difference” (p=0.016) and AUCSUM analyses (p=0.043).

Using AUCCHANGE, there were significant effects in favour of both treatment doses compared with placebo (p=0.036 and 0.016 for 44 µg and 22 µg IFNβ-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

**TOTAL COHORT ANALYSES WITH SCHEDULED VISIT DATA**

Utilising data from scheduled visits only, AUCSUM analysis on EDSS scores disclosed a trend favouring 22 µg IFNβ-1a over placebo (p=0.058). AUCCHANGE calculations showed that on EDSS scores, there were significant benefits of both doses over placebo (p=0.024 and 0.014 for 44 µg and 22 µg, respectively). For SNRS and AI data, treatment with IFNβ-1a 44 µg was also superior to placebo.

**Table 2.** Low (EDSS ≤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 AUCCHANGE for all visits: median (interquartile difference) and mean (SD)

<table>
<thead>
<tr>
<th>Entry EDSS ≤3.5 subgroup</th>
<th>AUC CHANGE (in EDSS-years)</th>
<th>Entry EDSS&gt;3.5 subgroup</th>
<th>AUC CHANGE (in EDSS-years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Two year difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFNβ-1a 44 µg (3x/week)</td>
<td>0 (1.0)</td>
<td>+0.21 (1.13)</td>
<td>0 (1.0)</td>
</tr>
<tr>
<td>IFNβ-1a 22 µg (3x/week)</td>
<td>0 (1.0)</td>
<td>+0.12 (1.27)</td>
<td>+0.5 (1.0)</td>
</tr>
<tr>
<td>Placebo</td>
<td>+0.5 (1.25)</td>
<td>+0.37 (1.18)</td>
<td>+0.5 (2.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+0.38 (1.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+0.64 (1.84)</td>
<td></td>
</tr>
<tr>
<td>Two year difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFNβ-1a 44 µg (3x/week)</td>
<td>−1.34 (7.66)</td>
<td>−0.12 (9.51)</td>
<td>−3 (16)</td>
</tr>
<tr>
<td>IFNβ-1a 22 µg (3x/week)</td>
<td>−1.07 (9.15)</td>
<td>−1.15 (11.49)</td>
<td>−2 (14)</td>
</tr>
<tr>
<td>Placebo</td>
<td>−2.49 (8.48)</td>
<td>−2.34 (11.22)</td>
<td>−4 (12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two year difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFNβ-1a 44 µg (3x/week)</td>
<td>0.19 (0.93)</td>
<td>0.16 (1.25)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>IFNβ-1a 22 µg (3x/week)</td>
<td>0.34 (1.13)</td>
<td>+0.41 (1.38)</td>
<td>+1 (1)</td>
</tr>
<tr>
<td>Placebo</td>
<td>0.29 (1.03)</td>
<td>+0.48 (1.44)</td>
<td>+1 (1)</td>
</tr>
</tbody>
</table>

AUCSUM analysis on EDSS scores disclosed a trend favouring 22 µg IFNβ-1a over placebo (p=0.058). AUCCHANGE calculations showed that on EDSS scores, there were significant benefits of both doses over placebo (p=0.024 and 0.014 for 44 µg and 22 µg, respectively). For SNRS and AI data, treatment with IFNβ-1a 44 µg was also superior to placebo.
Treatment comparisons between subgroups using data restricted to scheduled visits showed significant benefits of both treatment doses for patients with entry EDSS≤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, commonly employed end points associated with both disability and relapse assessment can be difficult to define and interpret. Meaningful clinical milestones, such as the conversion to secondary progression, 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, with the EDSS being the most criticized. Other scales (including SNRS and AI) have been less thoroughly evaluated. 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.

In this paper, we have analysed data from the PRISMS trial using two AUC methods, AUC<sub>SUM</sub> and AUC<sub>CHANGE</sub>. AUC<sub>SUM</sub> sums 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 large disability range in the cohort at baseline. On the other hand, AUC<sub>CHANGE</sub> obtained by normalising the AUC<sub>SUM</sub> value to the baseline rating is sensitive to in trial changes and has been previously utilised in a neurorehabilitation pilot study. 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<sub>SUM</sub> 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<sub>CHANGE</sub> 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. These patients have a tendency to deteriorate more rapidly and to be unresponsive to steroids and standard doses of IFNβ. The AUCCHANGE analysis of this subgroup disclosed that patients on placebo experienced strikingly more disability on EDSS, SNRS, and AI than their treated counterparts. AUCCHANGE confirmed the conventional end point of 2 year SNRS difference in showing significant treatment effects only for the 44 µg 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.

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, and the relatively lack of comparative data on SNRS and AI. 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 µg 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. 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, also needs to be considered. Although considered appropriate for analysing graded data, 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). 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 Florielle Dupont and Melvin Olsen for help in data analysis.

Appendix

AUCCHANGE: 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 (y0, y1, y2, ..., yn) are plotted versus their times of assessment (t0, t1, t2, ..., tn) totalling n+1 measurements y_i at times ti (i=0, ..., n), then

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

AUCCHANGE: This is the difference between the areas summated by AUC^{SUM} and the product of the baseline disability score y_i and the total time of study t_n, hence

\[ AUC^{CHANGE} = AUC^{SUM} - \frac{1}{2} y_0 t_n \]


