Multiple Sclerosis 2005; **11:** 626–634
www.multiplesclerosisjournal.com

Estimating long-term effects of disease-modifying drug therapy in multiple sclerosis patients

RA Rudick[*,1], GR Cutter[2], M Baier[3], B Weinstock-Guttman[4], MK Mass[5], E Fisher[6], DM Miller[1] and AW Sandrock[7]*

[1]Mellen Center for Multiple Sclerosis Treatment and Research, The Cleveland Clinic Foundation, Cleveland, OH, USA; [2]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA; [3]Center for Research Methods and Biometry, Cooper Institute, Denver, CO, USA; [4]Department of Neurology, Buffalo General Hospital, Buffalo, NY, USA; [5]Department of Neurology, Oregon Health Sciences University, Portland, OR, USA; [6]Department of Biomedical Engineering, The Cleveland Clinic Foundation, Cleveland, OH, USA; [7]Biogen Idec, Cambridge, MA, USA

Two methods were used to estimate the long-term impact of disease-modifying drug therapy (DMDT) in patients with relapsing multiple sclerosis (MS) who completed a placebo-controlled, randomized clinical trial of interferon beta-1a (IFNβ-1a). The study cohort consisted of patients with ambulatory relapsing MS who had previously participated in a placebo-controlled clinical trial for two years. At its end, patients were managed in an unstructured fashion by their neurologists and re-evaluated at an average of 6.1 years after the end of the trial. Follow-up evaluation was obtained for 93% of the 172 eligible patients. Because study inclusion criteria required that all patients have an Expanded Disability Status Scale (EDSS) score of $\leq$3.5 at entry, disability progression at follow-up was defined as EDSS $\geq$6.0. Two methods were used to estimate the expected proportions that reached EDSS $\geq$6.0 at follow-up. Estimates were compared with observed proportions. Method 1 used progression rates observed during the two-year phase III clinical trial and the percentage of time that patients were on DMDT during the follow-up period. Method 2 used progression rates from a natural history comparison group of relapsing–remitting MS patients. At the eight-year follow-up, 42.0% of the original placebo patients and 29.1% of the original IFNβ-1a patients reached an EDSS $\geq$6.0, an observed treatment effect of approximately 30%. Using method 1, it was estimated that 36.3% of the original placebo patients and 27.6% of the original IFNβ-1a patients should have reached an EDSS $\geq$6.0. Use of the natural history control group (method 2) predicted less plausible outcomes. Estimated proportions of patients reaching the endpoint were 63.3% for the original placebo group and 55.8% for the original IFNβ-1a group. Treatment effect sizes of 75–90% would be required to match estimates from method 2 with the observed outcome. The paucity of data on the long-term treatment of patients with MS may be aided by applying these or similar methods to vigorously followed cohorts of patients.
Multiple Sclerosis *(2005)* **11,** 626–634

Key words: *disease-modifying drug therapy; interferon beta; multiple sclerosis*

Introduction

Interferon beta (IFNβ) has been shown to reduce the relapse rate[1–3] and slow the progression of disability as measured by the Expanded Disability Status Scale (EDSS)[2–4] in patients with relapsing multiple sclerosis (MS). These benefits were demonstrated in short-term, randomized, controlled clinical trials. Critical information on long-term benefits of therapy is lacking; in particular, the long-term impact of IFNβ and other approved therapies on clinically meaningful disability progression is unknown.

**Correspondence: Richard A Rudick, Department of Neurology, Mellen Center for Multiple Sclerosis Treatment and Research, The Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA.*
E-mail: rudickr@ccf.org
Received 9 September 2004

One approach to this problem would be longer, randomized, placebo-controlled clinical trials; however, this is impractical for two main reasons. First, prolonged use of a placebo group is ethically questionable in patients with relapsing MS[5] because multiple, randomized, controlled clinical trials have demonstrated therapeutic benefits of drug therapy in these patients. In addition, the benefits of disease-modifying drug therapy (DMDT) may be most significant early in the disease.[6,7] Secondly, patients in a clinical trial are most likely to withdraw (drop out) from the trial if they are doing poorly. If the more severely affected patients are lost to follow-up, and this occurs more in one arm of the study than the other, the results are increasingly unreliable and potentially biased. Consequently, it is unlikely that long-term, randomized, placebo-controlled studies will ever be available to definitively assess the long-term impact of DMDT.

Various longitudinal clinical studies might be informative regarding the long-term impact of DMDT. For example, patients may be treated with standard therapeutic interventions in controlled, open-label studies[7,8] or may be contacted after some interval for standardized follow-

10.1191/1352458505ms1203oa

up after earlier participation in randomized, placebo-controlled or active therapy clinical trials. In these scenarios, however, there are no established methods to estimate the long-term benefits of the initial intervention.

This article describes two methods that can be used to estimate the long-term effects of controlled treatment with a DMDT, in this case illustrated using IFNβ-1a (Avonex, Biogen Idec, Inc., Cambridge, MA). The methods were applied to a cohort of patients from the initial IFNβ-1a placebo-controlled clinical trial who were randomized sufficiently early to complete two years of follow-up during the original phase III clinical trial.[2] These patients were located and invited to participate in standardized clinical assessments at an average of 8.1 years following entry into the original phase III clinical trial. Method 1 used the rate of progression to an EDSS score of ≥ 6.0 by the end of the original phase III clinical trial. Progression rates from the original phase III clinical trial were then applied to the follow-up period, and adjusted for months on DMDT during the uncontrolled follow-up interval. Method 2 used a natural history comparison group from Weinshenker *et al*. to estimate progression. [9] Various assumptions related to the methodology and resulting estimates of the long-term impact of DMDT on disease progression in MS are described herein.

Methods

This was a long-term follow-up study of patients who entered the randomized, double-blind, placebo-controlled, phase III trial of IFNβ-1a (Avonex)[2] sufficiently early to complete two years of follow-up in the original clinical trial. The primary objective of the study was to determine the relationship between Multiple Sclerosis Functional Composite (MSFC) scores during the controlled trial and status on EDSS, quality of life measures and magnetic resonance imaging (MRI) at follow-up. The study was reviewed and approved by the Institutional Review Board at each study centre. The study design and results from this follow-up study were reported.[10–14]

Patients

The original study used time to disability progression as the primary outcome; therefore, patients had variable lengths of follow-up based on when they were recruited. A total of 301 patients with relapsing MS, an EDSS score of 1.0–3.5, and an average disease duration of 6.5 years were randomized into the original phase III trial.[2] This cohort of patients (87 placebo, 85 IFNβ-1a) for whom disability progression rates at the two-year time point were available was eligible for the long-term follow-up study. Patients in this cohort were re-evaluated in the fall of 1999 by the investigators who conducted the original phase III clinical trial.

It is important to comment here on the type of selection used in identification of this study cohort of patients. Previous versions of this paper were criticized because only 57% of the total trial population survived for two years. We believe that this criticism is important to discuss because it may represent confusion in the MS community about the conduct of clinical trials.

In clinical trials, randomization is conducted to ensure unbiased allocation to study groups. Most trials stratify by clinical site and block their allocation, i.e., balance the allocation after short sequences to ensure that changes in recruitment over time are spread equally between treatment and control groups. Frequently, patients recruited early in a trial are different from patients recruited later. The first patients are usually from a 'prevalent' pool of patients, and the last patients are more 'incident' patients. How patients differ over time is usually unknown or unstudied. However, we are content with this process because randomization was balanced throughout the recruitment period and any differences occur equally in both groups.

Nevertheless, to make statements about an entire study, indeed, it is correct to be concerned with differences that arise over time. If all females are recruited first and then all males, one might be concerned with long-term statements being drawn only from earlier patients. However, this is a generalization issue rather than a biased or flawed analysis. The comparison of the earlier patients is still a valid comparison.

However, in a follow-up study, where only 57% of patients are re-examined, great concern over the selective nature of exactly which patients were examined should be expressed. This is known as informative censoring. The potential biasing effects of selective losses to follow-up are important to the validity of the analyses as well as their generalizability. The concern when a cohort that represents only 57% of the original cohort is identified sequentially at baseline is not the same concern. The randomization process makes this 57% a viable cohort to study prospectively, just as comparing treatments among males separately from treatments given to females, or comparisons of patients with MRI activity separately from those without MRI activity. To whom to generalize these results may be an important issue to discuss, but the comparisons themselves are valid.

Study procedures

An office visit was scheduled for each patient. When necessary, the visit was delayed at least eight weeks after treatment with corticosteroids. A detailed history of DMDT used after the end of the original phase III clinical trial was obtained from medical records and patient interviews. At study visits, patients were evaluated with the MSFC,[15] EDSS[4] and Sickness Impact Profile (SIP),[16,17] a validated measure of disease-related quality of life. Patients who were unable to travel to the clinic for evaluation completed the SIP and a self-administered version of the EDSS.[18] Seven patients died subsequent to the original phase III clinical trial but prior to the follow-up study (six from the placebo arm, one from the IFNβ-1a arm). In each case, relatives were interviewed and an EDSS was assigned based on the investigator's expert estimate of the patient's condition before death occurred in lieu of using an EDSS score of 10.

Primary endpoint

The primary endpoint for this study was the proportion of patients who had an EDSS ≥ 6.0 at the follow-up examination. This was considered clinically significant disability progression and is referred to as the endpoint in this paper. The first method used to estimate the proportion of patients reaching the endpoint was a form of direct adjustment of rates; it used the rates observed in the active and placebo arms during the original phase III clinical trial, adjusted the active arm rates for the duration of time on DMDT during the follow-up interval, and applied the placebo rates when not on therapy. The second method was a form of indirect adjustment for treatment because it used natural history control data to estimate the proportion of patients reaching the endpoint.

Estimating time on DMDT during follow-up

A detailed history of DMDT was obtained for each patient by chart review and interview. Treatment with IFNβ, glatiramer acetate, cyclophosphamide, mitoxantrone, azathioprine, methotrexate and corticosteroids was recorded, including start and stop dates. Duration of DMDT was calculated for each patient. Combinations of DMDT (e.g., IFNβ and methotrexate) were considered the same as use of a single DMDT. Time on therapy during the original phase III clinical trial was coded as 0 weeks for placebo patients and the number of weeks on study for IFNβ-1a patients. Time on therapy between the end of the original phase III clinical trial and the date of the follow-up examination was computed in weeks, and the total number of weeks of DMDT during this period was computed for each patient and summed. For example, patients on DMDT for 12 weeks, off for one year, on again for 12 weeks, then off for the duration of the follow-up period were assigned a total time on DMDT of 24 weeks. Similarly, a patient not on therapy for the duration of follow-up, except for the last 24 weeks, would also be assigned a total time on therapy of 24 weeks. The total time on DMDT was used to calculate the percentage of time on DMDT, i.e., treatment time divided by the total follow-up time.

Various assumptions are related to adjustments for open-label use of DMDT in this study. An important assumption was that different DMDT interventions had equivalent impact on EDSS progression, because no attempt was made to adjust for specific therapeutic interventions. A key assumption was that the drugs do not induce tolerance or permanently fix underlying pathology and, as such, only provide benefit while being taken. Another important assumption was that DMDT during the uncontrolled follow-up period had the same benefit, regardless of the timing of therapy and irrespective of whether therapy was provided over consecutive months or interrupted by periods off therapy. Another assumption was that combinations of drugs had the same impact as use of a single drug because no attempt was made to weight periods of DMDT for combinations of drugs.

Method 1: estimating outcome using rates observed during the phase III trial

The expected proportion of patients reaching the endpoint at follow-up was extrapolated from the rates observed at the two-year time point in the original phase III clinical trial. During the first two years, the probability of reaching the primary endpoint was 0.126 for placebo patients and 0.049 for IFNβ-1a patients. The probability that an individual patient would reach the endpoint during the follow-up period was computed using two scenarios. In the first, the expected values were generated for the entire eight-year duration, ignoring where the individual was at two years. In the second scenario, the estimate was computed projecting from year two to year eight and adding the expected number of new cases reaching the endpoint to that observed at two years.

An equation was used for calculating the probability of reaching the endpoint at follow-up, accounting for use of DMDT during the follow-up period. Three principal assumptions are related to this method: 1) the rate of progression to the endpoint was the same in each successive two-year interval as in the initial two-year interval; 2) the treatment effect for DMDT was the same in the follow-up period as during the two-year placebo-controlled interval; 3) each DMDT provided equivalent benefits, regardless of the timing of its use; 4) treatment is only effective while it is being given and there are no carryover effects. Summing the probabilities results in the estimated number of patients reaching the endpoint between two years and follow-up. To obtain the proportion expected to reach the endpoint over the follow-up period, the cumulative probabilities were corrected for the number of patients and added to the proportion found at the end of the phase III trial for each group. Further details are in the Appendix.

Method 2: estimating outcome using a natural history comparison group

The expected proportion of patients reaching the endpoint at follow-up was extrapolated from natural history data published by Weinshenker *et al*. in patients with relapsing MS.[9] Weinshenker *et al*. measured disability using the Disability Status Scale (DSS),[9] which consists of whole-step increments, rather than the EDSS, which consists of half-step increments, and this was taken into account. Different treatment-effect sizes from DMDT were then applied to adjust the estimates obtained from the natural history data. This enabled treatment-effect estimates to be analysed in terms of reasonableness of the estimates.

The equation used in this method for calculating the probability that an individual reached the endpoint at follow-up, given the observed EDSS at year two used three principal assumptions related to method 2: 1) for half-step increments of EDSS, the probability of transitioning was higher by half the distance between a certain transition and the integer DSS level; 2) for the duration of time at each step of the DSS, linear interpolation was used for the half-step increments to obtain times for the EDSS (i.e., if you started at 1.5, you would, on average, stay in that step for half the time that someone stayed if that patient were at

DSS = 1); 3) the natural history control group was an appropriate comparison group for the current study.

The natural history probabilities can be found in Weinshenker *et al*.[9] The expected number of patients reaching the endpoint of EDSS ≥6.0 was based on summing the probabilities of each patient reaching the endpoint, and dividing this number by the number of patients at follow-up, as follows: [Sum of adjusted *P* (EDSS ≥6.0)]/81 for the placebo group ($n = 81$) and [Sum of adjusted *P* (EDSS ≥6.0)]/79 for the IFNβ-1a group ($n = 79$). Further details related to this method are in the Appendix.

Results

Patients

Of 172 eligible patients, EDSS was determined at eight-year follow-up in 160 patients (93%). This included EDSS calculated at the time of a study visit in 137 patients, calculated from self-reporting in 16 patients, and assigned by the neurologist in seven patients who had died, as previously described.[13] Of the 160 patients with follow-up, 79 were from the original IFNβ-1a group and 81 were from the original placebo group.

Treatment with DMDTs

For the 160 patients with eight-year EDSS follow-up, the percentage of patients who were treated with various DMDTs after conclusion of the phase III clinical trial and the duration of treatment with DMDT are shown in Table 1. Of the 79 patients in the original IFNβ-1a cohort, 77 patients (97%) used IFNβ-1a during the follow-up interval compared with 52 of the 81 patients (64%) in the original placebo cohort. IFNβ-1b (Betaseron, Berlex Laboratories, Montville, NJ) was used by 25% of the original IFNβ-1a group and 31% of the placebo group. Many patients in the original phase III clinical trial used IFNβ-1b during the follow-up interval because IFNβ-1a was not available for approximately two years after the clinical trial ended. Other DMDTs were used but at a lower frequency. After completion of the study, the average number of weeks on IFN therapy (combined IFNβ-1a and IFNβ-1b) was 221 for the original IFNβ-1a group and 258 for the original placebo group.

Over the course of the follow-up interval, 29% of the original IFNβ-1a group and 28% of the original placebo group used at least two DMDTs. At the eight-year follow-up, 13.8% of the original placebo recipients were treatment-naïve. Excluding time during the original phase III clinical trial, patients from the original placebo group were on DMDT an average of 46.6% of the time whereas patients in the IFNβ-1a group were on a DMDT an average of 55.8% of the time.

Observed rate of disease progression

As per the inclusion criteria for the original phase III clinical trial, all patients had an EDSS ≤3.5 at entry, and therefore, no patients had an EDSS ≥6.0 at study entry. At the end of the original phase III clinical trial, 4.8% of patients in the IFNβ-1a group and 12.6% in the placebo group had progressed to an EDSS ≥6.0, representing a 62% reduction in progression to this endpoint with IFNβ-1a treatment in this cohort. Figure 1 shows that the benefits favouring the original IFNβ-1a group persisted at the eight-year follow-up visit. At eight-year follow-up, 29.1% of the original IFNβ-1a group reached the endpoint compared with 42.0% in the original placebo group ($P = 0.09$), representing a 31% reduction in progression to this endpoint with DMDT.

It is worth noting here that the 62% (1−4.8/12.6) reduction in progression to an EDSS ≥6.0 with IFNβ-1a at the end of the phase III trial is larger than the 37% reduction in disability progression as defined in the clinical trial (≥1.0 EDSS worsening, which required it be sustained for six months[2]). The effect on progression to EDSS ≥6.0 at the end of the phase III study could be exaggerated due to the small number of patients who reached this endpoint. At eight years' follow-up, 31% fewer patients in the original IFNβ-1a group reached an EDSS ≥6.0 compared with the original placebo group, a difference much closer to the results at two years using the original primary outcome measure.

Estimated rates of disease progression using methods 1 and 2

Table 2 shows estimated and observed progression rates using both scenarios of method 1. Both scenarios provided very similar estimates. For the scenario estimating rates over the entire eight years, the estimated progression rate at the eight-year follow-up was 26.2% for the IFNβ-1a group and 33.0% for the placebo group (a 20.6% reduc-

Table 1 Disease-modifying drugs used after the end of the clinical trial

Disease-modifying drug	*IFNβ-1a (n = 79)**		*Placebo (n = 81)***	
	No. of patients (%)	*Weeks, mean*	*No. of patients (%)*	*Weeks, mean*
IFNβ-1a	77 (97.5)	150	52 (64.2)	176
IFNβ-1b	20 (25.3)	71	25 (30.9)	82
Azathioprine	1 (1.3)	28	2 (2.5)	36
Adrenocorticotropic hormone	2 (2.5)	2	0	
Glatiramer acetate	4 (5.1)	56	8 (9.9)	22
Methotrexate	4 (5.1)	109	1 (1.2)	14

*IFNβ-1a patients were on a DMDT 46.6% of the time after the end of the clinical trial
**Placebo patients were on a DMDT 55.8% of the time after the end of the clinical trial

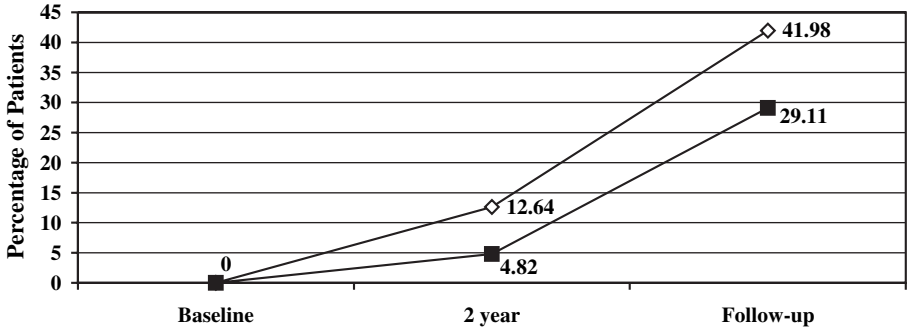

Figure 1 *Percentage of patients reaching an EDSS score ≥6.0 during phase III trial of IFNβa-1a.*

tion). Scenario 2, which projects forward from the observed two-year results, estimated a progression rate at the eight-year follow-up of 26.5% for the IFNβ-1a group and 33.8% for the original placebo group (a 21.6% reduction). The estimated rate in scenario 2 was computed for the placebo group as follows: 11 patients who had an EDSS ≥ 6.0 at two years plus patients expected to move to an EDSS ≥6.0 given that they were at an EDSS <6 at two years (expected cases = 16.73). Thus, the estimated probability of 0.338 was derived from (11 + 16.73)/(11 + 71), where the denominator consisted of the 11 patients with an EDSS ≥ 6.0 at two years plus 71 additional patients with outcome results at eight years. Similarly, the expected rate for reaching the outcome in the IFNβ-1a group was four patients with an EDSS ≥6.0 at two years plus 16.96 new cases expected. Thus, the estimated probability of 0.265 was derived from (4 + 16.96)/(4 + 75), where the denominator consisted of the number of original IFNβ-1a patients with an EDSS ≥6.0 at two years plus 75 additional patients with outcome results at eight years.

Table 3 shows estimated progression rates using method 2, assuming different treatment effect sizes. For these estimates to be in agreement with the observed proportion of patients reaching the endpoint, a treatment effect of 75–90% would be required.

Discussion

This paper demonstrates the use of two methods for estimating long-term effects of therapeutic intervention used in a short-term clinical trial in patients with relapsing MS. Method 1 produced estimates that closely matched the observed progression rates in the IFNβ-1a group. For example, progression estimates of 26.5% for the original IFNβ-1a group corresponded well to the observed progression rate of 29.1%. In contrast, it was estimated that approximately 33.5% of the original placebo group would reach the endpoint, whereas the observed progression rate was 42%. Thus, progression estimates for the original placebo group were approximately 10% higher than the observed outcome. One explanation for this observation, which remains conjectural, is that early treatment offers more advantage than does delayed treatment. We conjecture this because the underestimates that occur in both groups may be due to the model, but the underestimate of the placebo group is more than three times that of the IFNβ-1a group.

Using the natural history comparison group in method 2, estimated proportions of patients reaching the endpoint were 63.3% for the original placebo group and 55.8% for the original IFNβ-1a group, assuming no therapeutic benefit from treatment. Although the estimated progression rates were in parallel with observed differences at two years, these rates appear to have overestimated the treatment benefits during the entire follow-up interval. In order to get close to the estimates provided by method 2, treatment effect sizes of 75–90% would be required. However, based on clinical trial results showing treatment benefits on disability progression of approximately 30%,[2,3,19] it would seem unreasonable to expect longer-term effect sizes of 75–90%. Thus, it would seem that the natural history comparison group overestimated treatment effects; hence, studies that have used natural history data as a comparison group[20] should be interpreted with caution. This is consistent with the widespread observation that patients in natural history cohorts often do worse than those receiving placebo in clinical trial cohorts. Whether this results from selection bias related to selection for a clinical trial, placebo effects, changing patterns

Table 2 Method 1: proportion of patients with EDSS ≥6.0 at the follow-up examination

	Estimated over entire 8 years, %	*Estimated from 2 to 8 years, %*	*Observed*
IFNβa-1a	26.2	26.5	29.1
Placebo	33.0	33.8	42.0

Table 3 Method 2: proportion of patients with EDSS ≥6.0 at the follow-up examination

Treatment effect	*Placebo, %*	*IFNβ-1a, %*
None	63.3	55.8
50%	48.6	40.6
75%	41.3	33.1
90%	36.9	28.5
Observed proportion reaching the endpoint at follow-up	42.0	29.1

of medical care, changing natural history patterns or other factors is unknown. However, this result suggests that progression rates based on such natural history controls will overestimate event rates in clinical trials, possibly because of the use of the DSS compared with the EDSS or because of the changing face of patients with MS as diagnosis improves and time to diagnosis is shortened by MRI.

Because clinically relevant disability progression can be observed in a significant proportion of the population only after long observation periods, the methods presented here take advantage of a longer follow-up interval (8.1 years), during which evolution of the disease would be expected in a much larger proportion of patients than during a two-year clinical trial.[2] After the original phase III clinical trial ended, patients were treated with 'best medical practice' by their physicians and re-evaluated six years later. Thus, the current study population represents a real-world example of the problems encountered in conducting long-term follow-up, and this paper presents two methods of estimating expected progression that are necessary to assess the long-term impact of interventions. Although the methods are applied to the original IFNβ-1a study population, they are general approaches that could be applied to any long-term follow-up cohort, provided that the use of a DMDT can be accurately ascertained, and assuming near-complete follow-up to avoid the problem of informative censoring.

Finally, some controversy exists over whether this patient cohort is informative in long-term follow-up studies, an important issue in view of the publications related to this patient group.[10–14] The pivotal IFNβ-1a (Avonex) clinical trial was designed using survival analysis as the primary outcome, which was time to development of sustained EDSS worsening, into which 301 patients were randomized. The plan was for all patients to remain in the placebo-controlled clinical trial until the last randomized patient had completed two years' follow-up. However, because another form of IFNβ (Betaseron) was approved for use in July 1993, the IFNβ-1a study investigators raised concerns about the ethics of maintaining a placebo-controlled clinical trial when a similar product was available for clinical use following FDA approval. Additional concerns were expressed about potential loss of patients from the trial because of the availability of IFNβ in the clinics. These concerns were vetted with the NINDS, who largely sponsored the trial, and with the NINDS-appointed independent data and safety monitoring board (DSMB). A decision to terminate the trial prematurely was supported by both the study statistician and the independent statistician from the DSMB. The statisticians assured the investigators that the preplanned use of survival analysis permitted valid conclusions from the study. Furthermore, because the dropout rate was considerably below that projected, the study had already accumulated sufficient power to allow early termination. When the study was discontinued, 172 (57%) of the 301 patients had been in the study for at least two years, while 129 (43%) of the patients had been followed for less than two years.

In approximately 1999, the lead author of this article designed a long-term follow-up study to evaluate the significance of data collected during the clinical trial. A primary aim was to validate the MSFC by comparing two-year change in the MSFC to neurologic status at the long-term follow-up. Entry criteria for the long-term follow-up study included a requirement that patients had participated in the original clinical trial for at least two years, so that two-year MSFC data was available for each patient. The 172 patients who had been enrolled early enough in the trial were eligible for the follow-up study. We were able to evaluate 163 (95%) of these cases at an average of 8.1 years following randomization into the clinical trial. This cohort of 172 patients was entered sequentially, randomized to one of the two study arms, and followed under rigorous standardized conditions during the clinical trial. It is not a biased population, because only five patients were randomized early enough to have been potentially eligible for the long-term follow-up but dropped out of the study and thus are not included in the 172-patient cohort. Also, it is not a selected cohort – the 172-patient cohort was sequentially enrolled and was not selected from a larger pool of patients eligible for the follow-up study.

Even though the methods here are relatively simple, they provide for the examination of results using comparisons based on observed data. Although statistical techniques cannot adjust for selection biases, this study had a response rate of $>90\%$, affording the opportunity to explore methods of estimating without the great concern for informative censoring that occurs in most long-term open-label studies. Long-term follow-up studies that employ these methods should achieve an ascertainment rate of $>90\%$ to minimize concerns about informative censoring. As such, the methods presented herein have implications for current longitudinal clinical studies in patients with MS where the impact of DMDT must be estimated and taken into account.

Acknowledgements

The original clinical trial for this study was supported by Grant RO1 26321 from the National Institute of Neurological Disorders and Stroke, grant RG3099 from the National Multiple Sclerosis Society, and Biogen Idec, Inc. The long-term follow-up study was supported by Biogen Idec and by PO1 NS38667.

References

1 The IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing–remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* 1993; **43**: 655–61.
2 Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM *et al*. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Ann Neurol* 1996; **39**: 285–94.
3 PRISMS (Prevention of Relapses and Disability by Interferon β-1a Subcutaneously in Multiple Sclerosis) Study Group.

Randomised double-blind placebo-controlled study of interferon β-1a in relapsing/remitting multiple sclerosis. *Lancet* 1998; **352**: 1498–504.
4 Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; **33**: 1444–52.
5 Lublin FD, Reingold SC, and the National Multiple Sclerosis Society (USA) Task Force on Placebo-Controlled Clinical Trials in MS. Placebo-controlled clinical trials in multiple sclerosis: ethical considerations. *Ann Neurol* 2001; **49**: 677–81.
6 schwid SR, Bever CT Jr. The cost of delaying treatment in multiple sclerosis. What is lost is not regained. *Neurology* 2001; **56**: 1620.
7 The PRISMS (Prevention of Relapses and Disability by Interferon-β-1a Subcutaneously in Multiple Sclerosis) Study Group, and the University of British Columbia MS/MRI Analysis Group. PRISMS-4: long-term efficacy of interferon-β-1a in relapsing MS. *Neurology* 2001; **56**: 1628–36.
8 Herndon RM, Jacobs LD, Coats ME, Goodkin DE, Mass MK, Richert JR *et al*. Results of an ongoing, open-label, safety-extension study of interferon beta-1a (Avonex) treatment in multiple sclerosis. *Int J MS Care* [serial online] 1999; **1**: 1–6. Retrieved 5 May, 2003, from http://www.mscare.com
9 Weinshenker BG, Bass B, Rice GPA, Noseworthy J, Carriere W, Baskerville J, Ebers GC. The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. *Brain* 1989; **112**: 133–46.
10 Fisher E, Rudick RA, Cutter G, Baier M, Miller D, Weinstock-Guttman B *et al*. Relationship between brain atrophy and disability: an 8-year follow-up study of multiple sclerosis patients. *Mult Scler* 2000; **6**: 373–77.
11 Fisher E, Rudick RA, Simon JH, Cutter G, Baier M, Lee J-C *et al*. Eight-year follow-up study of brain atrophy in patients with MS. *Neurology* 2002; **59**: 1412–20.
12 Miller DM, Rudick RA, Baier M, Cutter G, Dougherty DS, Weinstock-Guttman B *et al*. Factors that predict health-related quality of life in patients with relapsing–remitting multiple sclerosis. *Mult Scler* 2003; **9**: 1–5.
13 Rudick RA, Cutter G, Baier M, Fisher E, Dougherty D, Weinstock-Guttman B *et al*. Use of the Multiple Sclerosis Functional Composite to predict disability in relapsing MS. *Neurology* 2001; **56**: 1324–30.
14 Rudick RA, Fisher E, Lee JC, Duda JT, Simon J. Brain atrophy in relapsing multiple sclerosis: relationship to relapses, EDSS, and treatment with interferon β-1a. *Mult Scler* 2000; **6**: 365–72.
15 Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fischer J *et al*. Recommendations from the National Multiple Sclerosis Society Clinical Outcomes Assessment Task Force. *Ann Neurol* 1997; **42**: 379–82.
16 Bergner M, Bobbitt RA, Kressel S, Pollard WE, Gilson BS, Morris JR. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *Int J Health Serv* 1976; **6**: 393–415.
17 John Hopkins University. *The Sickness Impact Profile User's Manual and Interpretation Guide*. Baltimore, MD: Johns Hopkins University, Department of Health Policy and Management, 1996.
18 Ratzker PK, Feldman JM, Scheinberg LC, Larocca NG, Smith CR, Giesser BS, Aisen ML. Self-assessment of neurologic impairment in multiple sclerosis. *J Neurol Rehab* 1997; **11**: 207–11.
19 Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP *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.
20 Johnson KP, Brooks BR, Ford CC, Goodman A, Guarnaccia J, Lisak RP *et al*. Sustained clinical benefits of glatiramer acetate in relapsing multiple sclerosis patients observed for 6 years. *Mult Scler* 2000; **6**: 255–66.

Appendix A
Equations for calculating the probability of EDSS ≥6.0 at follow-up

Method 1: estimating outcome using rates observed during the phase III trial

The probabilities of reaching the endpoint assumed that the rates that would apply during the follow-up period were the same as those observed during the first two years (105 weeks). Follow-up time was calculated individually for each patient and used to determine the number of two-year periods from which to apply these probabilities. That is, we calculated the probability of how many two-year intervals the patient had to experience without reaching the endpoint and then subtracted this from one, obtaining the probability of reaching the endpoint at some time during the follow-up period. For example, if the probability of reaching the endpoint was 0.12 in two years, then the probability of not reaching the endpoint in four years would be $(0.88) \times (0.88)$, or 0.774, because there are two two-year intervals during follow-up. Consequently, the probability of reaching the endpoint at some time over the four-year period would be $1.00 - 0.774 = 0.226$.

This approach assumed that an individual would remain at an EDSS ≥6.0 after reaching the endpoint. Given the variability in the EDSS scores, we know this is not technically correct for the data observed. For example, four of the 14 patients who had follow-up examinations at eight years were at an EDSS ≥6.0 at two years and an EDSS <6 at follow-up (three of 10 in the placebo group, one of four in the IFNβ-1a group). This represents a source of error in the predictions.

$$\text{Prob6} = \text{pctonrx} \times (1 - (0.951807)^{(\text{fuperiod}/105)}) + (1 - \text{pctonrx}) \times (1 - (0.873563)^{(\text{fuperiod}/105)})$$

where prob6 is the probability of reaching the endpoint EDSS ≥6, pctonrx is percentage of time on treatment, 1 − pctonrx is percentage of time off treatment, 0.951807 is the probability of not reaching the endpoint during DMDT, based on IFNβ-1a treatment, 0.873563 is the probability of not reaching the endpoint during placebo treatment and fuperiod/105 is the follow-up period in weeks divided by

105 weeks of original study to yield number of periods of follow-up time.

Method 2: estimating outcome using a natural history comparison group

To estimate the proportion of patients reaching a given EDSS level based on natural history observations on the DSS, a multistep process was followed for each patient. Given the starting EDSS, the probability that a patient would move through the remaining EDSS levels, reaching a DSS ≥ 6 over the time period of follow-up, was computed for each patient. To obtain the expected number of patients reaching EDSS, these probabilities were summed. Because a person who starts at a DSS = 1 must transition from a DSS = 1 to a DSS = 2, then from 2 to 3, and so on, within the time period, the probability was lower for patients who started lower on the EDSS scale. This process assumes that progression through each subsequent stage of the DSS is independent of the previous transition. Because the natural history control data contained DSS rather than EDSS information, we estimated progression probabilities for EDSS levels by assuming the time was halved and the chances of progressing were increased compared with the DSS scales. That is, for the half-step increments of the EDSS, the probability of transitioning to the next step was assumed to be the same as the integer DSS level plus 50% of the difference between a 1.0-point change and the previous DSS integer level (i.e., the probability of transition to the next DSS step is more likely because a patient is 'closer' to the next step). For duration of time between steps of the DSS, linear interpolation was used for the half-step increments in the EDSS (i.e., a patient starting at 1.5 would transition to 2.0 in half the time required for a patient to move from 1.0 to 2.0). For example, if the probability of moving from DSS = 2 to DSS = 3 is 0.81, then the probability of moving from EDSS = 2.5 to EDSS = 3.0 would be increased to $0.81 + [(1 - 0.81) \times 50\%] = 0.905$. Furthermore, if the median time to transition from EDSS = 2 to EDSS = 3 was 3.2 years, the time to move from 2.5 to 3 would be halved, or 1.6 years.

The EDSS at the end of the original phase III clinical trial was used as the starting point from which to calculate the probability of reaching the endpoint by the eight-year follow-up visit because the two-year results were known. The published median time at each DSS level was used to calculate the expected time for transition from one DSS level to the next. For each patient, the expected times for each transition from one DSS level to another were summed, starting with the end-of-trial EDSS and proceeding to an EDSS = 6.0. For example, based on natural history data, a patient at an EDSS = 3 at the end of the original phase III clinical trial would require 6.30 years to reach the endpoint. This is based on the sum of 2.17 (median time in years to move from DSS = 3 to DSS = 4), 2.07 (median time in years to move from DSS = 4 to DSS = 5) and 2.06 (median time in years to move from DSS = 5 to DSS = 6). Because the amount of follow-up time varied slightly for each patient, an adjustment was made to estimate the fraction of 6.30 years of exposure that a patient would have experienced. That fraction K was used to compute the probability of not reaching the endpoint during the follow-up period. The estimated probability of not reaching the endpoint was calculated, and the result was subtracted from one to indicate the probability of achieving the endpoint at some time during the follow-up period.

The following is an example of calculating the probability of reaching the endpoint for a patient at EDSS = 3.0 at the end of the original phase III clinical trial, subsequently followed for six years:

$$K = 6/(2.17 + 2.07 + 2.06)$$
$$= 6/6.3$$
$$= 0.95$$

[This indicates that it took 95% of the usual time to reach an EDSS ≥ 6 given that observation of the patient began at an EDSS = 3.0 and the patient was followed for six years divided by the natural history time to go through the three DSS steps (DSS = 3, 4 and 5) to get to a DSS = 6, which is, on average, 6.3 years.]

The probability of reaching an EDSS = 6 over this period is one minus the probability of not reaching it, as follows:

$$\text{Prob} = 1 - \{1 - (0.82 \times 0.88 \times 0.94)\}^{0.95}$$
$$= 1 - \{1 - 0.678304\}^{0.95}$$
$$= 1 - \{0.321696\}^{0.95}$$
$$= 1 - 0.340466 = 0.659534$$

This probability was then adjusted for the time on DMDT during the follow-up period (i.e., if therapy reduced this probability by x%, then what is the percentage of patients reaching an EDSS = 6).

$$\text{Adjusted } P(\text{EDSS} \geq 6.0)$$
$$= (1 - \%\text{ time on Rx}) \times P(\text{EDSS} \geq 6.0) + (\%\text{ time on Rx}) \times (1 - \text{Rx effect size}) \times P(\text{EDSS} \geq 6.0)$$

where

$$\text{Adjusted } P(\text{EDSS} \geq 6.0)$$
$$= \text{probability adjusted for time on DMDT}$$

$$\%\text{ time on Rx}$$
$$= \text{time on DMDTduring follow-up/time of follow-up}$$

$$P(\text{EDSS} \geq 6.0)$$
$$= \{1 - [1 - (\text{Natural history probability of reaching endpoint})^K]\},$$

where

K = (follow-up)/natural history time required to reach the outcome

Rx effect size = estimated treatment–effect size (% reduction in progression due to Rx)

For a patient at an EDSS = 3.0 after the phase III trial, $P(\text{EDSS} \geq 6.0)$ is $\{1-[1-(0.82 \times 0.88 \times 0.94)^{\text{follow-up period}/(2.17+2.07+2.06)}]\}$, where 0.82 is the natural history probability of moving from an EDSS = 3.0 to an EDSS = 4.0, and so on, and 2.17 is the natural history median time it takes to move from EDSS = 3.0 to EDSS = 4.0, and so on.