**Predictive validity of NEDA in the 16- and 21-year follow-up from the pivotal trial of interferon beta-1b**

Douglas S. Goodin, MD;1 Anthony T. Reder, MD;2 Anthony L. Traboulsee, MD;3 David K. B. Li, MD;3 Dawn Langdon, PhD;4 Gary Cutter, PhD;5 Stuart Cook, MD;6 Timothy O’Donnell, DO;7 Marcelo Kremenchutzky, MD8 Joel Oger, MD9 Ralf Koelbach,10,a Christoph Pohl, MD11,12† Eva-Maria Wicklein, MD11 for the IFNB Multiple Sclerosis Study Group and the 16- and 21-Year LTF Investigatorsb

1Department of Neurology, University of California, San Francisco, California, United States

2Department of Neurology, University of Chicago, Chicago, Illinois, United States

3University of British Columbia, Vancouver, British Columbia, Canada

4Department of Psychology, Royal Holloway, University of London, London, UK

5Department of Biostatistics, UAB School of Public Health, Birmingham, Alabama, USA

6Department of Neurosciences, Rutgers University, Newark, New Jersey, USA

7Pompton Lakes Pulmonary P.C., Lincoln Park, New Jersey, USA

8Western University, London Health Sciences Centre, London, Ontario, Canada

9Department of Neurology, University of British Columbia, Vancouver, British Columbia, Canada

10PAREXEL International, Berlin, Germany

11Bayer AG, Berlin, Germany

12University Hospital of Bonn, Bonn, Germany

†Deceased

aPerformed statistical analyses

bMembers listed in the E-Appendix

**Corresponding author**

Douglas S. Goodin

Department of Neurology

University of California, San Francisco

350 Parnassus Ave, #908

San Francisco, CA 94117

415-514-1684

douglas.goodin@ucsf.edu

**Key words:** multiple sclerosis; autoimmune diseases, intereron beta-1b, prognosis

**Abstract**

**Background:** Long-term follow-up from the randomized-trial of interferon beta-1b (IFNB-1b) permitted the assessment of different definitions of no-evidence-of-disease-activity (NEDA) for predicting long-term outcome in MS.

**Objective:** To examine the predictive-validity of different NEDA definitions.

**Methods:** Predictive-validity for negative-disability outcomes (NDOs) at 16 years and survival at 21 years post-randomization were assessed. NEDA in the first 2 years was defined as, clinical-NEDA: no relapses or EDSS-progression from baseline to Year 2; NEDA-3a: no relapses, no confirmed ≥1-point EDSS-progression, and no new T2-active lesions; NEDA-3b: no relapses, no EDSS-progression, and no increase in T2 burden-of-disease (T2-BOD); and NEDA-4: no relapses, no EDSS-progression, and no increase in T2-BOD or atrophy. NDOs were defined as death, need for wheelchair, EDSS ≥6, or progressive MS.

**Results:** 245 and 371 patients were evaluated at 16 and 21 years, respectively. Clinical-NEDA predicted NDOs (*P*=0.0029), as did baseline EDSS (*P*<0.0001), baseline T2-BOD (*P*<0.0001), and change in T2-BOD (*P*=0.0033). IFNB-1b treatment (*P*=.0251), relapse rate in the 2-years before study start (*P*=0.0260), T2-BOD at baseline (*P*=0.0014), and change in T2-BOD (*P*=0.0129) predicted survival at 21 years.

**Conclusions:** Clinical NEDA predicted long-term disability outcome. By contrast, definitions of NEDA that included on-therapy changes in MRI variables did not increase the predictive-validity.

**Introduction**

No-evidence-of-disease-activity (NEDA) has become a popular therapeutic target and an endpoint in clinical trials of new disease-modifying therapies (DMTs) for patients with multiple sclerosis (MS).1 In addition to the presumed benefits on quality of life for patients who achieve NEDA status early in the course of their disease, this measure may also predict long-term disease outcomes. For example, one study found that NEDA two years after the start of treatment predicted disability seven years after study start.2 By contrast, another recent study assessing NEDA in a non-randomized cohort followed for 10 years found that NEDA was unrelated to outcome.3

Understanding the value of NEDA measured over a short time-period requires validation of its relationship to long-term patient outcomes. The most appropriate long-term datasets for this purpose are follow-up of patients from randomized, placebo-controlled clinical trials (RCTs) because RCTs consist of very well-characterized patient populations who have been followed closely for multiple years with patients who have been divided at the start of the trial into two or more comparable groups which differ only in their exposure to active therapy. The original trial of interferon beta-1b (IFNB-1b) for patients with active relapsing-remitting MS (RRMS) offers the most extensive dataset available, which matches these specific criteria.4-6 IFNB-1b-treated patients in the trial had reduced rates of disease activity (both relapses and MRI changes) compared with placebo.5,6 Patients initially randomized to receive IFNB-1b during this RCT also showed better clinical outcomes and substantially improved survival 16 and 21 years (respectively) following randomization.7-10 Thus, the long-term follow up from the original trial of IFNB-1b offers a unique opportunity to assess the predictive value of NEDA for outcomes after an extended period of treatment and follow-up.

NEDA is usually defined as no relapses, no disability progression, and no new or enlarging lesions on MRI (NEDA-3).1 Although this is the most common definition of NEDA, the inclusion or exclusion of various parameters defining disease activity have also been explored. Thus, in addition to NEDA-3, measures of brain atrophy can also be included in the definition of NEDA, a circumstance which is commonly referred to as NEDA-4.8 Brain volume loss is thought to provide a more comprehensive assessment of disease status and because brain volume loss correlates with disability progression and cognitive impairment.11 Nevertheless, the current lack of widespread implementation of brain atrophy measurements may limit the feasibility of incorporating NEDA-4 into clinical practice. Simpler definitions of NEDA have also been proposed that focus only on relapses and disability progression (excluding MRI measures of disease activity or atrophy),2 referred to here as clinical-NEDA.

The objective of the present analysis is to examine the predictive value of NEDA for negative disability outcomes (NDOs) and death in the patients from the long-term follow-up of the original RCT of IFNB-1b in RRMS.5 Multiple definitions of NEDA were examined to assess the relative impact of various clinical and MRI components in regression models.

**Materials and methods**

*Study design*

Full details of the methods used for the pivotal IFNB-1b study have been published previously.5 Briefly, patients with RRMS and no previous exposure to immunosuppressive treatment were randomly assigned to receive alternate-day IFNB-1b (50 µg or 250 µg) or placebo. Requirements for study entry included the occurrence of ≥2 relapses in the two years prior to randomization and an Expanded Disability Status Scale (EDSS) score ≤5.5. Following the RCT, after a maximum of 5.5 years, patients in the placebo-arm were offered open-label treatment with IFNB-1b 250 µg. All procedures were approved by the institutional review boards of the participating institutions and patients provided informed consent at enrollment.

Long-term outcomes were assessed in this population at 16 and 21 years post-randomization. At 16 years, the principal investigators at all of the 11 original centers attempted to arrange in-person evaluations of patients enrolled to determine each patient’s medical history subsequent to the RCT, including MS-related symptoms, disease status (RRMS or SPMS), disability, treatment, MRI, and adverse events.12 In addition, a follow-up using public database information to ascertain survival was conducted 21 years after the original randomization.10 The present analysis assessed the occurrence of NDOs at 16 years (death, need for a wheelchair, progression to unremitting EDSS ≥6.0, or progression to secondary progressive MS after Year 2), and survival at 21 years.

The purpose of combining these different outcomes into a single composite variable (NDO) was three-fold. First, the inclusion of several outcomes increased the total number of observed NDOs (thereby increasing the statistical power of the study). Second, each element of the NDO composite represents a so-called “hard” or “unequivocal” negative disability endpoint.11 And finally, analysis of the 16-year data had previously demonstrated that the impact of therapy on each of these endpoints (considered either individually or collectively) was substantially similar.9

In the present study NEDA was defined in 4 ways: clinical NEDA (no relapses and no EDSS progression from baseline to Year 2), NEDA-3a (no relapses, no confirmed ≥1-point EDSS progression by Year 2, and no new T2-active lesions since baseline during the first 2 years of treatment), NEDA-3b (no relapses, no EDSS progression from baseline to Year 2, and no increase in T2 burden-of-disease (BOD) from baseline to Year 2 (ie, no enlargement in the volume of hyper-intense lesions on T2-weighted images – measured in cubic centimeters), and NEDA-4 (no relapses, no EDSS progression from baseline to Year 2, no increase in T2-BOD from baseline to Year 2, and no increase in brain atrophy from baseline to Year 2 (ie, no enlargement of the 3rd ventricle size). Gadolinium (Gd) was not administered as part of this study. Nevertheless, it is important to recognize that the occurrence of Gd-enhancement only reflects current disease activity, which has taken place over a window of three months or less. By contrast, the measurement of new T2-activity over 2 years is an estimate (although probably an underestimate) of the total disease activity, which has taken place over the entire 2-year period.

For the purposes of this study, as in the original report from this RCT,6 new MRI activity (evaluated blindly) was defined as: 1) new lesions (those that had never been seen before); 2) recurrent lesions (lesions appearing at the same site where a previous lesion had disappeared); and 3) an enlarging lesion (>70% increase for small [≤1 cm] lesions and >10% for large [>1 cm] lesions). Third ventricular width (in mm) was measured manually (and in a blinded fashion) by one of the investigators (ALT) at the University of British Columbia.

*Statistical analyses*

Unifactorial (single factor) logistic regression analyses were conducted using the following covariates: clinical-NEDA, NEDA-3a, NEDA-3b, and NEDA-4, EDSS at baseline, sex, age (at onset of symptoms and at start of study), time since onset of clinical symptoms, annualized relapse rate (ARR, rate of relapses per year) before start of treatment and from baseline to Year 2, EDSS change from baseline to Year 2, confirmed 1-point EDSS progression within the first two years, Multiple Sclerosis Severity Scale (MSSS) score at RCT start,T2-BOD at baseline, T2 activity (cumulative number of new, enlarging, or recurrent lesions) at Year 2, change in T2-BOD from baseline to Year 2, 3rd ventricle size at baseline, change in 3rd ventricle size from baseline to Year 2, initial treatment assignment in the RCT (with both doses of IFNB-1b considered together), and the presence of neutralizing antibodies (NAbs) during the RCT.

Multifactorial analyses using stepwise elimination procedures (p=0.5 to enter and p>0.1 to exit) were conducted including clinical-NEDA, NEDA-3a, NEDA-3b, or NEDA-4 and EDSS at baseline, sex, age (at onset of clinical symptoms and start of study), time since onset of clinical symptoms, ARR before start of treatment and at Year 2, T2 activity at Year 2, T2-BOD at baseline, change in T2-BOD from baseline to Year 2, 3rd ventricle size at baseline, change in 3rd ventricle size from baseline to Year 2, initial treatment assignment, and the occurrence of NAbs during the RCT. Because of overlap with the composite measures themselves, the on-study relapse rate and progression measures were excluded from the multifactorial analyses of all NEDA definitions, the change in T2-BOD from baseline was additionally excluded from the analyses of NEDA-3b and NEDA-4, new T2 activity at Year 2 was excluded from the analyses of NEDA-3a, and the change in 3rd ventricle size from baseline was excluded from the analyses of NEDA-4.

**Results**

*16-Year results*

Following the RCT (N=376), 245 patients had NDO assessments at 16 years (Table 1). NDOs were present in 129 patients (52.7%) in the 16-year population. Patients with NDOs by NEDA status are shown in Table 2a. As expected, more patients achieved NEDA when MRI measures were excluded from the definition (n=43 with clinical-NEDA alone, compared to n<20 using the other definitions).

At 16 years, clinical-NEDA (as shown in Figure 1) predicted the likelihood of experiencing NDOs in both unifactorial (OR 0.421 [95% CI, 0.212, 0.836], *P*=0.0135) and multifactorial models (OR 0.159 [95% CI, 0.047, 0.534], *P*=0.0029). In addition, baseline EDSS, time since symptom onset, T2-BOD at baseline, MSSS at baseline, 3rd ventricle size at baseline, change in EDSS from baseline to last RCT visit after 2 years, ARR over the 2 years from the start of the RCT, and 3rd ventricle size at Year 2 all significantly predicted the probability of NDOs at Year 16 in unifactorial models (Figure 1).

By contrast, NEDA-3a, NEDA-3b, and NEDA-4 did not predict the likelihood of NDOs either in unifactorial or in multifactorial models (Figure 1). Nevertheless, in these multifactorial models, there were certain covariates that seemed to be consistently associated with the likelihood of NDOs: EDSS at baseline, T2-BOD at baseline, and change in T2-BOD from baseline to Year 2 (clinical-NEDA and NEDA-3a only).

*21-Year results*

Survival was assessed in 98.7% (371/376) of patients at 21 years after the RCT (Table 1). Deaths by NEDA classification are shown in Table 2b. Unlike in the 16-year analysis, we did not find an association between clinical-NEDA and survival at 21 years. Nevertheless, there was a trend for such an association in both the unifactorial (HR 0.485 [95% CI, 0.223, 1.054], *P*=0.0678) and multifactorial models (HR 0.359 [95% CI, 0.111, 1.159]; *P*=0.0865). By contrast, except for NEDA-3a (*P*=0.0927), there was not even a suggestion of such an association for the other definitions of NEDA (Figure 2). However, in the unifactorial models, T2-BOD at baseline, initial treatment with IFNB-1b, change in EDSS from baseline to last visit after 2 years, confirmed 1-point EDSS progression at Year 2, ≥1-point change in EDSS from baseline to Year 2, ARR over the 2 years after the start of IFNB-1b, new T2 lesions at Year 2, and 3rd ventricle size at Year 2 were all significantly associated with the likelihood of death after 21 years. And, in multifactorial models, the covariates consistently associated with the probability of survival were the initial randomization to treatment with IFNB-1b and T2-BOD at baseline.

**Discussion**

The pivotal trial of IFNB-1b for patients with RRMS provides the most extensive and complete dataset available that matches each of the criteria for assessing the long-term predictive value of NEDA outlined above. This RCT demonstrated that IFNB-1b-treated patients had reduced rates of both relapses and MRI activity compared with placebo. 4-6 Patients initially randomized to receive IFNB-1b also showed better clinical outcomes and substantially improved survival 16 and 21 years (respectively) following randomization.7-10 Thus, the long-term follow up from the original trial of IFNB-1b offers a unique opportunity to assess the predictive value of NEDA for outcomes after an extended treatment and follow-up period.

The present analysis showed that patients who experienced clinical-NEDA for the 2-year period following randomization in the RCT were less likely to develop NDOs after 16 years. There was also a trend for them to have better survival. However, when MRI parameters (new T2 lesions, T2-BOD, or 3rd ventricle size) were included within the definition of NEDA, this composite no longer significantly predicted the probability of these outcomes. Consequently, our findings suggest that although clinical-NEDA may be an important therapeutic target for patients on DMTs, the incorporation of MRI parameters into the definition diminishes the predictive validity of NEDA.

Although our findings suggest that the inclusion of MRI measures into the definition of NEDA lessens its predictive value, it is important to recognize that the number of patients achieving clinical-NEDA was 2 to 7 times higher than the number achieving NEDA by other definitions. On the one hand, this means that our study had less statistical power to test the hypothesis that MRI measures actually add to the predictive value of clinical NEDA. On the other hand, there is nothing in the point estimates for these alternative NEDA definitions to suggest that the incorporation of MRI added anything to clinical-NEDA and, in some instances, their addition seemed to subtract (Figures 1 and 2). Additional studies with much larger patient numbers may be better able to answer fully this question. However, at present, there are no other trials that have comparable duration and completeness of follow-up compared with the present study.

Our NDO results at 16 years are similar to the findings of Rotstein et al.,2 who examined the predictive value of NEDA-3a for disability outcomes in 215 patients with RRMS or clinically isolated syndrome from 2000 to 2005. Both studies showed a predictive value for NEDA early in treatment. Importantly, their study showed that NEDA has a predictive value of disability outcomes after 7 years; our study demonstrated a value of clinical-NEDA over a much longer time interval (16 years).

None of the NEDA definitions significantly predicted survival at 21 years, although there were trends in this direction for clinical-NEDA and NEDA-3a. This finding cannot be ascribed simply to the small number of patients who achieved NEDA status because, if anything, there were more patients who reached NEDA status for the 21-year analysis than there were for the 16-year analysis despite the fact that clinical-NEDA was a significant predictor of the occurrence of NDOs at 16 years. We demonstrated previously that some (~20%) of the deaths which occurred in this cohort were due to non-MS causes and many others were due to the consequences of so-called “end-stage” MS, which many patients never reach in their lifetimes.10 Perhaps, such a dilution-effect or selection bias may have adversely impacted the statistical significance of the observed association. This seeming paradox will require further study to resolve.

Despite the fact that incorporating MRI measures of activity and/or severity into the definition of NEDA did not improve its predictive capacity, certain individual MRI measures (eg, T2-BOD at baseline) were significant predictors in both unifactorial and multifactorial regression analyses at 16 and 21 years. This indicates that there is clearly a role for assessment of outcome measures not included in the composite NEDA endpoint. Nevertheless, it should be noted that many of these measures (eg, baseline MRI or clinical findings) cannot be used to judge the success or failure of therapy. This result is consistent with other studies showing a relationship between T2 lesion volume for long-term disability outcomes in patients with RRMS.13-17 One such study found that the predictive capacity improved by combining measures of T2 lesion volume and brain atrophy. This particular finding differs from those of the present study, at least insofar as these two measures were components of the NEDA composite.

In addition, cognition is an early independent predictor of physical disability progression over 10 years18 and this additional clinical manifestation of MS may be a more reliable indicator than silent pathological changes. It also appears that cognitive decline may occur in the context of NEDA-3a,19 suggesting cognition should be monitored to capture disease impact more comprehensively.19,20 Additional large studies are needed to investigate this hypothesis.

This study has certain potential limitations including the retrospective nature of the analysis, the use of a relatively small patient population, the fact that at the start of the RCT, other than having MS, patients were generally healthy, the fact that after the RCT many patients received alternative therapies to IFNB-1b,12 and the fact that MRI techniques have advanced considerably since pivotal study of IFNB-1b was begun. Some of these potential limitations are less critical than others. For example, the retrospective aspects of this study are substantially mitigated by the fact that all of the baseline and on-study measures used to determine association with outcome were obtained contemporaneously and prospectively. Moreover, the 16-year outcomes included in the NDO composite were so-called “hard” (ie, fixed and unequivocal) disability endpoints and, in any event, these outcomes were each determined without any knowledge of how the patient had fared during the RCT. Similarly, in the 21-year analysis, the outcome of interest was death, which is an unambiguous outcome that cannot give rise to any concerns about the adequacy of blinding. In addition, with 98.7% case-ascertainment in the 21-year study, there can also be no concerns about ascertainment bias. Thus, any concerns about the “retrospective” nature of the analysis are misplaced. With regard to health status at the start of the RCT, this certainly could have contributed to an increased likelihood of survival in the next 20 years. Nevertheless, any bias should have affected all patient-subgroups equally and, in any circumstance, this could not explain the increased survival in patients exposed to IFNB-1b during the RCT.10 Moreover, the time-course of survival in placebo-randomized patients seemed to be superimposable with the population-based survival course of MS reported elsewhere.10 Thus, it seems unlikely that this supposed “healthier” status of our cohort influenced our current findings. The fact that some patients received alternative therapies to IFNB-1b,12 is a common occurrence in all long-term follow-up studies and, indeed, this trial is distinguished by the fact that, for 8 years after the start of the RCT, no therapeutic alternatives to IFNB-1b were available and at 16 years, over half (55%) of the patients had exclusively received IFNB-1b.12 More importantly, however, it is difficult to rationalize how the receipt of alternative therapies could have biased any analysis of NEDA’s predictive validity.

By contrast, concerns about sample size are more pertinent. There is no question that the numbers of patients achieving NEDA in this study were relatively small, especially as more parameters were included in the definition of NEDA. This reduced the statistical power to detect the associations of NEDA with outcome. Nonetheless, there are several reasons to believe that this concern is also misplaced. Certainly, it is true that many modern RCTs have up to 3 times the patient numbers that the IFNB-1b trial had. However, none of these RCTs has the duration of placebo-controlled exposure and none has anywhere near either the long-term follow-up or the level of case-ascertainment as our study. Thus, the IFNB-1b trial represents a unique dataset, in which patients were followed as randomized, blinded cohorts for a period of 5.5-years and were subsequently followed for an additional 16 years with essentially complete ascertainment (98.7% of the original study cohort for survival analyses). Moreover, the lack of other effective DMT options other than IFNB-1b prior to 1996 means that the patients with MS who participated in this trial were only exposed to either IFNB-1b or no DMT for a period of 8-10 years following randomization. In addition, although the statistical power may have been low to detect a significant association of NEDA-3a, NEDA-3b, or NEDA-4 with outcome, the fact remains that the point estimates for these associations suggest that these alternative definitions of NEDA added nothing to that already provided by clinical-NEDA and in some cases these alternative definitions seemed to detract (Figures 1 and 2). And finally, rather than being attributable to a small sample size, the failure of these NEDA definitions, which included MRI, to predict outcome is more likely to be a reflection of the fact that these MRI measures are not correlated with disability outcome – either when they are considered by themselves or when they are included, together with other variables, in a step-wise regression analysis.21

Similarly, concerns about the presumed superiority of “modern MRI techniques” compared with those that were in use in 1993 may be valid. For example, at the time of the IFNB-1b trial, the available MRI techniques were limited to the measurement of new, enlarging or recurrent T2 lesions; total BOD; and 3rd ventricular width. Today, by contrast, the use of both Gd-enhancement and fluid attenuated inversion recovery (FLAIR) imaging have become standard MRI methods. Moreover, we are now better able to quantify atrophy (including the use of segmentation to assess grey and white matter separately), to achieve better spatial resolution using high field strength magnetic fields (eg, 7 T), to provide high-quality images of the spinal cord, to detect more (and more persistent) disease activity using triple dose Gd administration, to use diffusion tensor imaging to gage disruption of neural pathways, and to determine the local environment of water molecules using magnetic transfer imaging. Ultimately, the use of these new techniques may make MRI a more useful tool for predicting outcomes in MS. However, the caveat to this is that these new techniques will need to be validated in the real world of patients with MS.

In addition, these older MRI methods (eg, the measurement of new T2 lesion-activity, the assessment of brain atrophy as measured by 3rd ventricular width, and the determination of T2 disease-burden) are still in wide-spread clinical and research use. It is, therefore, noteworthy that these older MRI techniques seem not to contribute much (if anything) to the prediction of outcome. Whether newer MRI techniques will be shown, ultimately, to make an important contribution to outcome-prediction remains an open and unanswered question. Finally, it is worth noting that, if anything, the use of some modern MRI techniques (eg, FLAIR imaging, Gd-enhancement, thinner slice thicknesses, and improved spatial resolution) will only serve to reduce the number of patients who achieve NEDA-3a and NEDA-3b status at any timepoint compared with our study.

In summary, our results demonstrate that clinical-NEDA (defined by the lack of both relapses and EDSS progression) predicted NDOs at 16 years, underscoring the value of clinical measures for assessing outcome both in MS clinical trials and in MS patients individually. By contrast, despite MRI measures of T2-BOD and brain atrophy consistently predicting long-term outcomes, their inclusion as variables in alternative definitions of NEDA (eg, NEDA-3a, NEDA-3b, and NEDA-4) did not increase the predictive capacity of clinical-NEDA by itself.

**Acknowledgments**

This study was funded by Bayer HealthCare Pharmaceuticals. We are grateful to the patients and the investigators for their continuing contributions to the study. The authors would also like to thank Dr George C. Ebers (University Department of Clinical Neurology, John Radcliffe Hospital, Oxford, UK) for his contributions to the design of the long-term follow up study. Robert C. Ristuccia, PhD (Precept Medical Communications) provided medical writing assistance (editing of the manuscript, preparation of figures and tables) that was funded by Bayer HealthCare Pharmaceuticals.

**Declaration of conflicts of interest**

* DSG has received either personal compensation (for consulting, serving on a scientific advisory board, or speaking) or financial support for scholarly activities from pharmaceutical companies that develop products for MS, including Bayer-Schering Pharma, Merck-Serono, Teva Pharmaceuticals, Questcor/Malinkrodt, and Novartis
* ATR has received either personal compensation (for consulting, serving on a scientific advisory board, or speaking) or financial support for scholarly activities from the American Medical Association, Institute for Health Care Quality, NARCOMS, and National Multiple Sclerosis Society, and from pharmaceutical companies that develop products for and other entities, including Abbott Laboratories, Astra Merck, Athena Neurosciences, Aventis Pharma, Bayer Schering Pharma, Berlex Laboratories, Biogen and Biogen Idec, BioMS Medical Corp., Blue Cross-Blue Shield, Caremark Rx, Genentech, Genzyme Corporation, GlaxoSmithKline, Hoechst Marion Roussel Canada Research, Hoffman-LaRoche, Idec, Immunex, Neurocrine Biosciences, Novartis Corporation, Parke-Davis, Pfizer, Pharmacia and Upjohn, Questcor/Malinkrodt, Quintiles, Serono, Sandoz (now Novartis), SmithKline Beecham, Takeda Pharmaceuticals, and Teva-Marion
* ALT reports no conflicts of interest.
* DKBL has received research funding from the Canadian Institute of Health Research and Multiple Sclerosis Society of Canada. He is the Director of the UBC MS/MRI Research Group which has been contracted to perform central analysis of MRI scans for therapeutic trials with Genzyme, Hoffmann-LaRoche, Merck-Serono, Nuron Biotech, Parexel, and Sanofi-Aventis. He has also acted as a consultant to Vertex Pharmaceuticals and served on the Data and Safety Advisory Board for Opexa Therapeutics and Scientific Advisory Boards for Novartis, Nuron, and Roche.
* DL has received payment for consultancy services from Novartis, Bayer, TEVA, and Biogen; served on speakers bureaus for Almirall, TEVA, Biogen, Novartis, Bayer, and Excemed; and received research grants from Novartis, Biogen, and Bayer. All are paid to DL’s university.
* GC has received either personal compensation (for consulting, serving on a scientific advisory board, or speaking) or financial support for scholarly activities from pharmaceutical companies that develop products for MS and other entities, including Antisense Therapeutics Limited, Sanofi-Aventis, Bayhill Pharmaceuticals, BioMS Pharmaceuticals, Daichi-Sankyo, Genmab Biopharmaceuticals, Glaxo Smith Klein, PTC Therapeutics, Medivation, Eli Lilly, Teva, Vivus, University of Pennsylvania, NHLBI, NINDS, NMSS, Ono Pharmaceuticals, Alexion Inc., Accentia, Bayer, Bayhill, Barofold, CibaVision, Novartis, Diagenix, Consortium of MS Centers, Klein-Buendel Incorporated, Enzo Pharmaceuticals, Peptimmune, Somnus Pharmaceuticals, Teva pharmaceuticals, UTSouthwestern, Visioneering Technologies, Sandoz, and Nuron Biotech, Inc.
* SC has received personal compensation for consultations or lectures from Merck-Serono, Bayer HealthCare, Sanofi-Aventis, Neurology Reviews, Biogen Idec, Teva Pharmaceuticals, Genmab and Actinobac Biomed Inc. He has served on scientific advisory boards for Bayer HealthCare, Merck-Serono, Actinobac Biomed, Teva Pharmaceuticals, and Biogen Idec.
* TO has received compensation from Bayer for sitting on the adjudication committee for this study and has participated as an investigator for open-label, and safety and efficacy trials with Otsuka, Schering, and Bristol-Myers Squibb. He has served on data safety monitoring boards for MedImmune, Nabriva, and Moderna and also on clinical adjudication committees for GSK and Merck.
* MK’s institution receives research funding from the Canadian Institute for Health Research (CIHR) and the Chapman Chair in MS Clinical Research from the London Health Sciences Centre Foundation; he received compensation for scholarly activities from Biogen, Genzyme and Novartis; he participated as site investigator for clinical trials from Biogen, Sanofi-Genzyme, Novartis, EMD Serono, Teva, and Bayer HealthCare Pharmaceuticals.
* JO has received honoraria, consulting fees, travel, research and/or educational grants from Aspreva, Aventis, Bayer, Biogen Idec, EMD Serono, Genentech, Schering, Talecris, and Teva Neurosciences.
* RK is a salaried employee of PAREXEL International
* CP was a salaried employee of Bayer AG
* EMW is a salaried employee of Bayer AG

**References**

1. Giovannoni G, Turner B, Gnanapavan S, Offiah C, Schmierer K, Marta M. Is it time to target no evident disease activity (NEDA) in multiple sclerosis? *Mult Scler Relat Disord.* 2015;4(4):329-333.

2. Rotstein DL, Healy BC, Malik MT, Chitnis T, Weiner HL. Evaluation of no evidence of disease activity in a 7-year longitudinal multiple sclerosis cohort. *JAMA Neurol.* 2015;72(2):152-158.

3. Cree BA, Gourraud PA, Oksenberg JR, et al. Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol.* 2016;80(4):499-510.

4. The IFNB Multiple Sclerosis Study Group. Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology.* 1995;45(7):1277-1285.

5. 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(4):655-661.

6. The IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group. *Neurology.* 1993;43(4):662-667.

7. Ebers GC, Reder AT, Traboulsee A, et al. Long-term follow-up of the original interferon-beta1b trial in multiple sclerosis: design and lessons from a 16-year observational study. *Clin Ther.* 2009;31(8):1724-1736.

8. Goodin DS, Ebers GC, Cutter G, et al. Cause of death in MS: long-term follow-up of a randomised cohort, 21 years after the start of the pivotal IFNbeta-1b study. *BMJ Open.* 2012;2(6).

9. Goodin DS, Jones J, Li D, et al. Establishing long-term efficacy in chronic disease: use of recursive partitioning and propensity score adjustment to estimate outcome in MS. *PLoS One.* 2011;6(11):e22444.

10. Goodin DS, Reder AT, Ebers GC, et al. Survival in MS: A randomized cohort study 21 years after the start of the pivotal IFNbeta-1b trial. *Neurology.* 2012;78(17):1315-1322.

11. Kappos L, De SN, Freedman MS, et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. *Mult Scler.* 2016;22(10):1297-1305.

12. Ebers GC, Traboulsee A, Li D, et al. Analysis of clinical outcomes according to original treatment groups 16 years after the pivotal IFNB-1b trial. *J Neurol Neurosurg Psychiatry.* 2010;81(8):907-912.

13. Kearney H, Rocca MA, Valsasina P, et al. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. *Mult Scler.* 2014;20(1):72-80.

14. Popescu V, Agosta F, Hulst HE, et al. Brain atrophy and lesion load predict long term disability in multiple sclerosis. *J Neurol Neurosurg Psychiatry.* 2013.

15. Radue EW, Barkhof F, Kappos L, et al. Correlation between brain volume loss and clinical and MRI outcomes in multiple sclerosis. *Neurology.* 2015;84(8):784-793.

16. Rudick RA, Lee JC, Simon J, Fisher E. Significance of T2 lesions in multiple sclerosis: A 13-year longitudinal study. *Ann Neurol.* 2006;60(2):236-242.

17. Traboulsee AL, Cornelisse feminine P, Sandberg-Wollheim M, et al. Prognostic factors for long-term outcomes in relapsing-remitting multiple sclerosis. *Mult Scler J Exp Transl Clin.* 2016;2:2055217316666406.

18. Moccia M, Lanzillo R, Palladino R, et al. Cognitive impairment at diagnosis predicts 10-year multiple sclerosis progression. *Mult Scler.* 2016;22(5):659-667.

19. Damasceno A, Damasceno BP, Cendes F. No evidence of disease activity in multiple sclerosis: Implications on cognition and brain atrophy. *Mult Scler.* 2016;22(1):64-72.

20. Portaccio E, Stromillo ML, Goretti B, et al. Neuropsychological and MRI measures predict short-term evolution in benign multiple sclerosis. *Neurology.* 2009;73(7):498-503.

21. Goodin DS, Traboulsee A, Knappertz V, et al. Relationship between early clinical characteristics and long term disability outcomes: 16 year cohort study (follow-up) of the pivotal interferon b-1b trial in multiple sclerosis. *J Neurol Neurosurg Pshcyiatr*. 2012;83(3):282-287.

**Table 1.** Patient characteristics

|  |  |  |
| --- | --- | --- |
|  | **16-year analysis (N=245)** | **21-year analysis (N=371)** |
| Female, n (%) | 169 (69.0) | 259 (69.8) |
| Age (years) at onset, mean (SD) | 27.5 (6.8) | 27.4 (6.9) |
| Age (years) at start of treatment, mean (SD) | 35.2 (7.3) | 35.5 (7.1) |
| EDSS at baseline, median (range) | 2.5 (0.0-5.5) | 3.0 (0-5.5) |
| Duration of disease (years) from onset, mean (SD) | 7.74 (6.06) | 8.05 (6.19) |
| Initial treatment with IFNB-1b, n (%) | 168 (68.6) | 249 (67.1) |
| Baseline MSSS, mean (SD) | 4.11 (2.22) | 4.32 (2.25) |
| T2-BOD (cm2) at baseline, mean (SD) | 19.73 (20.91) | 20.68 (21.37) |
| Ventricle size (mm) at baseline, mean (SD) | 4.750 (2.032) | 4.966 (2.323) |
| ARR in the 2 years prior to study start, mean (SD) | 1.66 (0.75) | 1.68 (0.80) |
| Change in EDSS from baseline to Year 2, mean (SD) | -0.17 (1.15) | 0.11 (1.40) |
| Confirmed ≥1-point EDSS progression at Year 2, n (%) | 38 (15.5) | 86 (23.2) |
| EDSS ≥1 point above baseline at Year 2, n (%) | 45 (18.5) | 98 (26.6) |
| ARR 2 years after start of IFNB-1b, mean (SD) | 1.13 (1.28) | 1.26 (1.40) |
| Change in T2-BOD (cm2) from baseline, mean (SD) | 1.16 (7.45) | 1.53 (7.51) |
| T2 activity at Year 2, mean (SD) | 2.53 (3.37) | 2.57 (3.49) |
| Ventricle size (mm) at Year 2, mean (SD) | 5.479 (2.316) | 5.669 (2.537) |
| Change in ventricle size (mm) from baseline to Year 2, mean (SD) | 0.683 (1.056) | 0.645 (1.030) |
| NAb positive, n (%) | 74 (30.2) | 109 (29.4) |

ARR, annualized relapse rate (rate of relapses per year); BOD, burden of disease; EDSS, Expanded Disability Status Scale; IFNB-1b, interferon beta-1b; MSSS, Multiple Sclerosis Severity Scale; NAb, neutralizing antibodies.

**Table 2.** NDOs after 16 years (a) and deaths at 21 years (a), n/N (%)a

|  |  |  |
| --- | --- | --- |
|  | **Patients achieving NEDA** | **Patients not achieving NEDA** |
| 1. **NDOs at Year 16**
 |
| Clinical-NEDA | 15/43 (34.9) | 112/200 (56.0) |
| NEDA-3a | 7/19 (36.8) | 105/196 (53.6) |
| NEDA-3b | 5/15 (33.3) | 100/189 (52.9) |
| NEDA-4 | 2/7 (28.6) | 78/161 (48.4) |
|  |
| 1. **Death at Year 21**
 |
| Clinical-NEDA | 7/59 (12) | 72/310 (23) |
| NEDA-3a | 2/27 (7) | 68/304 (22) |
| NEDA-3b | 2/19 (11) | 61/280 (22) |
| NEDA-4 | 1/9 (11) | 54/245 (22) |

NEDA, no evidence of disease activity.

aOnly patients with non-missing data for NEDA variables included.

**Figure Captions**

**Figure 1.** Unifactorial (a) and multifactorial (b) predictors of NDOs 16 Years after randomization

ARR, annualized relapse rate; BOD, burden of disease; EDSS, Expanded Disability Status Scale; IFNB-1b, interferon beta-1b; MSSS, Multiple Sclerosis Severity Score; NAb, neutralizing antibody; NDO, negative disability outcome; NEDA, no evidence of disease activity.

**Figure caption:** Results of unifactorial (a) and step-wise multifactorial (b) models for NDOs 16 years after randomization. Significant predictors in Figure 1a are shown in red. Figure 1b shows only those covariates which were statistically significant predictors in multifactor models. NEDA-3a, NEDA-3b, and NEDA-4 were not significantly associated with either NDOs or survival and did not enter into the final multifactorial regression equations.

**Figure 2.** Unifactorial (a) and multifactorial (b) predictors of death 21 years after randomization

ARR, annualized relapse rate; BOD, burden of disease; EDSS, Expanded Disability Status Scale; IFNB-1b, interferon beta-1b; MSSS, Multiple Sclerosis Severity Score; NAb, neutralizing antibody; NEDA, no evidence of disease activity.

**Figure caption:** Results of unifactorial (a) and step-wise multifactorial (b) models for survival 21 years after randomization. Significant predictors in Figure 2a are shown in red. Figure 2b shows only those covariates which were statistically significant predictors in multifactor models. NEDA-3a, NEDA-3b, and NEDA-4 were not significantly associated with either NDOs or survival and didn’t enter into the final multifactorial regression equations.







