Ebers, Heigenhauser, Daumer, Lederer, Noseworthy:

Multiple sclerosis, the measurement of disability and access to clinical trial data

Sonderforschungsbereich 386, Paper 429 (2005)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Multiple sclerosis, the measurement of disability and access to clinical trial data

GC Ebers, L Heigenhauser, M Daumer, C Lederer, and JH Noseworthy

Department of Clinical Neurology, University of Oxford, Radcliffe Infirmary, Oxford OX2 6HE, UK (Prof GC Ebers); **Sylvia Lawry Centre for Multiple Sclerosis Research, Munich, Germany** (L Heigenhauser, Dr M Daumer, Dr C Lederer); **Department of Neurology, the Mayo Clinic College of Medicine, Rochester, Minnesota** (Prof JH Noseworthy)

Correspondence to: Prof GC Ebers, Department of Clinical Neurology, University of Oxford, Radcliffe Infirmary, Oxford OX2 6HE george.ebers@clinical-neurology.oxford.ac.uk

Summary

Background Inferences about long-term effects of therapies in multiple sclerosis (MS) have been based on surrogate markers studied in short-term trials. Nevertheless, MS trials have been getting steadily shorter despite the lack of a consensus definition for the most important clinical outcome - unremitting progression of disability.

Methods We have examined widely used surrogate markers of disability progression in MS within a unique database of individual patient data from the placebo arms of 31 randomised clinical trials.

Findings Definitions of treatment failure used in secondary progressive MS trials include much change unrelated to the target of unremitting disability. In relapsing-remitting MS, disability progression by treatment failure definitions was no more likely than similarly defined improvement for these disability surrogates. Existing definitions of disease progression in relapsing-remitting trials encompass random variation, measurement error and remitting relapses and appear not to measure unremitting disability.

Interpretation Clinical surrogates of unremitting disability used in relapsing -remitting trials cannot be validated. Trials have been too short and/or degrees of disability change too small to evaluate unremitting disability outcomes. Important implications for trial design and reinterpretation of existing trial results have emerged long after regulatory approval and widespread use of therapies in MS, highlighting the necessity of having primary trial data in the public domain.

Introduction

The natural history of multiple sclerosis (MS) evolves over some 30-40 years[1]. The most important therapeutic target is unremitting disability occurring in the progressive phase of the disease. A secondary progressive phase (hereafter SPMS) supervenes in >80% of relapsing-remitting patients and after 15-18 years, some 50% of patients need assistance to walk, are confined to wheelchair, bed or have died[2]. Maintaining randomisation/placebo arms beyond 2-3 years in individuals of reproductive age has been difficult. In this condition where therapeutic disappointment is familiar, regular reports have appeared from short-term trials showing treatment effects on surrogate markers[3]. These studies highlight problems common to many chronic diseases in extrapolating from short-term surrogates.

Unremitting disability is the outcome most relevant yet least clearly measured in relapsing-remitting MS (hereafter RRMS). Relapses, short-term disability change, and magnetic resonance imaging (MRI) are surrogates for this. No consensus definition of disability accumulation can be found in trials leading to approved therapies. On a 10-point scale (EDSS - expanded disability status scale) unconfirmed and confirmed (by repeated serial in-trial observations - see below) changes of 0.5 and 1.0 point have been used as primary and secondary outcomes in studies with results leading to regulatory approval. Longer term data have been unavailable[4,5] or dropouts have compromised interpretation[6,7]. Meanwhile, dependence on inferences from short-term measures of relapses, “disability” and from MRI-based surrogate markers has continued both for pilot and phase 2/3 studies. SPMS is characterised by more advanced EDSS levels where variation is less than in RRMS[8]. However, the course of SPMS has been largely intractable to therapy[9,10,11]. The dearth of long-term outcome data stimulated this study, aimed at identifying reliable markers or surrogates of long-term disease progression.

Materials and methods

Centre description

The Sylvia Lawry Centre for Multiple Sclerosis Research (hereafter SLC) was established in February 2001 to support independent research into MS natural history and clinical trial methodology in order to accelerate development of effective therapies. This unique database consisted of the placebo arms from 31 treatment trials. See **http://www.slcmsr.info** for further information.

Dataset

Analyses are based on the assembled SLCMSR data set of RR and SP patients from 31 trials. This is split into open (40% -1344 patients), closed (50%) and reserve (10%) parts,

the latter two as a reservoir for confirmation of results. Clinical trials were of one to five years' duration (most <2-3 years) with planned 3-6 monthly assessments.

Patient selection

We defined 3 data subsets:

Subset 1- RR and SP MS patients observed for at least 15 months, having uninterrupted serial assessments every three months (± 1 month) and with EDSS at baseline below 6.0 -. This was necessary in order to compare different definitions of 'time to sustained progression'.

Subset 2 had one observation two years (± 1 month) after entry into the study but without any further restrictions.

Subset 3 comprised those who were thought to have measurements taken during relapses.

The characteristics of the subsets are listed in Table 1, by RR and SPMS phenotype. From the original 1344 placebo patients in the open dataset, 425 patients entered subset 1 and 516 entered subset 2. Truncating data to simulate a 2-year trial minimises the impact of dropouts and exclusions which rapidly accelerate after this duration.

Replication

This was carried out for the most stringent definition of progression (1 point confirmed at 6 months) in an independent, randomly selected, matched group in the closed dataset.

Outcome criteria

These were the definitions of unconfirmed or confirmed disability change of 0.5-1.0 EDSS points used in the database trials. Confirmation times of 3-6 months, added to the 3-month minimal period for a change to register, shorten the effective period of observation. Few studies required additional visits after study conclusion, so the last on-study visit could only be used for confirmation.

Definitions of worsening (trial defined 'treatment failure' or TF) and placebo 'treatment improvement' (TI)

Using the EDSS scale, we examined: 1) 0.5 points minimum increase unconfirmed/confirmed, 2) 1.0 minimum increase unconfirmed/confirmed. EDSS changes sustained at 90 days or 180 days as in the trials the change were deemed *confirmed*. Improvement was a *decrease* in the same defined scores.

Clinical phenotype definitions

This was defined as in the RR and SP trials constituting the dataset.

Sensitivity of these measures

Trial-employed definitions of disability worsening have been widely accepted for clinical decision-making. To test the validity of these definitions of worsening or 'treatment failure', we examined placebo arms from the 31 trials, generally reported to have been stable for at least 1-3 months at baseline. We reasoned that the difference between probabilities of significant improvement vs significant worsening would approximate the proportion of worsening attributable to noise from random variation/measurement error. Data were also analysed omitting the first point, effectively extending the effective period of prestudy stability to > 4-6 months.

Methods of analysis

For perspective on within-trial changes for EDSS, we calculated events meeting the selected outcome criteria. Proportions of opposite events with same EDSS change and confirmation period were calculated. Since MS trials typically have used Kaplan-Meier curves to describe the accumulation of disability, these were plotted for definitions of worsening and improvement used in the trials studied and p-values from a two-sided logrank test were computed. Additionally, one-sided Wilcoxon signed rank tests were used to compare absolute differences in EDSS over one and two years.

Results

Clinical characteristics of the 3 database subsets are in Table 1 showing representative characteristics of trial-eligible RR and SP patients.

Table 1 near here if possible on left facing page along with Table 2 and 3.

Defined EDSS change by time in study

We calculated worsening (meeting treatment failure definitions) and improvement (same definition but opposite polarity) and counted events in RR and SP for each of the 6 periods of confirmation/non confirmation using all available data for subset 1 without truncation - shown in Table 2. In SP, worsening occurred more frequently than improvement and was less and less likely to remit with increased stringency of definition. However, even in SP, improvement occurred 53% as often as worsening (Table 2). For RR however, event frequencies were similar for all definitions and sustainability was also similar for less stringent definitions. The longer the confirmation period, the fewer were the events and overall, the lower the likelihood of reversion, but only a trend is seen in the direction of fewer reversions for the more stringent definitions of worsening/improvement.

Table 3 shows the distribution by magnitude/confirmation of EDSS changes in SP and RR subgroups. In SP, worsening >improvement but a substantial proportion of worsening is offset by improvement events. However for RR, worsening was not significantly more likely than improvement, even for 2 year definitions in subsets 1 and 3.Only for the 2-year definition in subset 2 with incomplete 3-month data points was marginal significance reached but not surviving correction for multiple comparisons $p<0.07$. The Wilcoxon was not significant for this subgroup. The data for SP contrast with RR as there are many more individuals who have worsened for all degrees of change and the number of improvements is proportionately less indicating the diminished measurement noise in SP.

Figs 1 and 2 near here

Survival curves for SP are seen in Figs 1A-F for each of the 6 definitions of worsening/improvement listed in the Fig. legend, illustrating that sustained worsening is more likely than similarly defined sustained improvement. Table 4 shows that for all definitions of SP, worsening is significantly more likely than improvement using survival analysis (Kaplan-Meier). Nevertheless, difference between sustained worsening and improvement rivals that between baseline and improvement.

The RR survival curves in Figs 2 A-F do not significantly distinguish between worsening and improvement. The high frequency of "sustained improvement" in RRMS is seen for all definitions. By 2 years, half of placebo arm patients would have a 0.5 pt. decrease confirmed at 3 months, only slightly less often than for sustained progression. Overall, the probabilities of 0.5 and 1.0 point changes at the 3-monthly intervals up to 2 years are remarkably similar in the K-M curves for sustained "improvement" vs sustained worsening. For RR and SP patients, we truncated the data at 2 years to simulate a 2-year trial, but no significant difference for RR is seen using all data points (Table 4). K-M analysis assumes independence of the curves, which is not strictly true, and we are comparing different events in the same patients, but potential overlap between patients with worsening/improvement is limited by time constraints imposed by their definitions and by trial durations.

We asked if findings were an artefact of the EDSS levels reached. We hypothesised that if treatment failure definitions represented random noise/measurement error, we could run the analysis in reverse, taking the last EDSS point as the first and the first as last. No difference in the relative survival curves for improvement and worsening was found (not shown). This further suggests that variation derived from rater and subject unrelated to true disability level was measured.

Replication from the closed data set for the 1 point 6 month confirmed change in EDSS in subset 1 showed again no significant difference between sustained worsening vs improvement (p=0.184).

Discussion

The problems of therapy evaluation in MS are mostly familiar. A chronic disease with substantial variation in short and long-term outcome, MS has proven formidably challenging. Recently, several therapies have been shown to favourably affect relapse-related clinical and MRI outcomes[7,12]. These surrogates for unremitting disability accumulation, the primary medical and economic concern in MS, have heavily influenced therapeutic decisions, making validation of outcomes used in trials overdue. An independent data resource aimed at facilitating the discovery of effective treatments made such validation analyses possible. The SLC database size facilitates generalizability of findings, retesting of secondary hypotheses and replication.

The disability scale (EDSS), ubiquitous in these trials, has clinical relevance at higher levels but contains weaknesses. Some variation derives from inconsistent subject performance within the symptomatic spectrum of disease. Furthermore, progressive MS cannot be confidently diagnosed in most patients until an EDSS of 4.0 on a 0-10 point scale is reached. At this and higher levels, there is reasonably good agreement on scores of ≥4. However for disability levels characterising trials claiming effectiveness in RRMS, interrater variation is one point or greater 40% of the time[8]. Nevertheless, 0.5 or 1 point changes[7,13,14] have defined treatment failure for Kaplan-Meier curves in several pivotal trials, conflicting with recommendations from the designer of the scale[15] and with general considerations about precision and accuracy. Measurements of changes smaller than the variation intrinsic to the tool/object are considered inappropriate at best.

Some patients meeting criteria for treatment failure in these studies must do so by random variation/measurement error. For quantification, we compared worsening to improvement, each defined by magnitude and duration of +/- confirmed EDSS change in-

trial. These results support EDSS-dependent definitions of treatment failure used in progressive MS trials, but the degree of variation as measured by "improvement" in placebo-treated progressive cases was large. To the extent that sustained improvement is a measure of noise or measurement error unrelated to unremitting progression, more than half of treatment failure events in placebo-treated progressive cases are offset by "improvement" events. Accepting random variation/measurement error as treatment failure surely diminishes power of studies to detect effectiveness and increases vulnerability of study conclusions to the influence of imperfect blinding[16]. Measurement accuracy varies inversely with both variance and the *square* of bias.

RRMS analyses contrast with the SP results. Although greater change on the scale and longer confirmation intervals were seen to increase specificity of treatment failure definitions, no clinical measure of disability we evaluated can be supported as measuring unremitting disability in RRMS. Changes of 0.5 points are unambiguously invalid, even confirmed at 3 or 6 months. Similarly, 1 point changes were not significantly more likely to occur for worsening than for improvement, although a trend appears in the expected direction. Survival curves were almost superimposable for all definitions and worsening/improvement KM comparisons showed no statistical significance. The usual course of RRMS is eventually manifested by sustained upward movement on the scale, but 2 years in trial-eligible patients may be insufficient to show it. Selection for frequent relapses and against progressive disease may well bias against sustained short-term disability change.

We considered that the findings reflected ascertainment idiosyncrasies of RR trials. In these, pre-trial disease stability was usually required for 1-3 months, duration insufficient to eliminate those who would improve spontaneously from recent relapse within the early trial period. However, after excluding the first data point and effectively extending stability out to 4 - 6+ months for most cases, the relative survival curves for improvement and worsening were essentially unaltered. Furthermore removal of values taken during identified relapses had no impact on findings and conclusions (not shown).

The pre-analysis division of the overall data set into two components (open/closed), each separately and serially analysed, allowed for replication of results. Natural history studies show that unremitting change requires an elapsed year for confirmation[17] and in-trial times less than this have been associated with a high rate of spontaneous reversion[2,18,19]. The findings do seem to mirror the degree of interrater variability[8,20] and have major implications for future trial design. Our own recommendation for strengthening SP outcome criteria underestimated what is required[21].

If "treatment failure" occurs via random variation/measurement error or remitting relapses but is misinterpreted or misrepresented as unremitting disability, progressive erosion of study power results. Kaplan-Meier analysis, ubiquitous in MS trials and appropriate for hard endpoints, is unsuitable for trials where outcomes are highly susceptible to random variation, measurement error and relapses. Relapses are already counted in trials as independent outcomes. It will require more careful assessment of relapse and MRI surrogates to put clinical outcomes in proper context. Meanwhile, definitions of treatment failure/unremitting disability change require reconsideration. Clinical decisions in RRMS dependent on these outcome measures must be seen as outside an evidence base, notwithstanding the approval of the largest selling interferon by the US FDA for the indication of preventing disability using measures evaluated here[14]. Since worsening was no more frequent than improvement over 1-2 years in placebo arms

of RRMS trials, debate regarding the propriety of placebo arms should not be concluded, pending validation of relapse and MRI surrogates[22]. For the average MS patient having 0.5 attacks /year, 6 years of treatment equates to one attack prevented unless there is concomitant reduction of disability. This has to be shown not assumed.

The role of academic investigators and regulatory agencies in RRMS studies may also warrant review. Opponents are reminded of the wide acceptance a decade ago of the disability outcomes these studies could not validate[23]**.** Which disability outcomes should be relied on in future MS trials? Increasing the degree of disability change and the duration of confirmation would increase power in SP studies and seems essential for meaningful disability results in RRMS. Despite oft-repeated claims that EDSS disability measures are insensitive to change[24], reducing measurement error, accounting for unblinding and extending the duration of trials may be more formidable obstacles to valid conclusions. However, trial sample sizes in RRMS have successively enlarged while duration has progressively shortened since the first pivotal interferon study[12]**.** A recent FDA decision rested on surrogate markers and p-values in *12-month* data[25], the published clinical results originally only 6 months. In contrast, we have had difficulty in this study showing more worsening than improvement in *two-year* placebo data.

We have not had access to treatment arm data but similar limitations for disability outcomes are probable. These data do not contradict well-documented, short-term reductions of relapse rate and MRI T2 reported in RRMS trials[21]. Similar validation steps for these surrogate measures are in process at the SLC.

Contributors

Martin Daumer is the director of the Sylvia Lawry Centre and coordinated data accrual.. This study was conceived by CL, MD and GE. The analyses were done by CL and LH. GE and JN (Chairman) are on the SLC Scientific Oversight Committee. GE wrote the first draft of the manuscript and coordinated the final version with input from all authors.

Conflict of interest

There is no conflict of interest identified.

Acknowledgements

We gratefully thank other participants in the SLC, the public-spirited companies and individuals who generously donated their data on placebo arms with the expressed view of improving clinical trial methodology, anonymous private donors and several national MS Societies. We also acknowledge useful advice from colleagues especially Albrecht Neiss, George Loudon, Christine Purdy, Sarah Phillips, Ingrid Kreuzmair for supporting the statistical analysis. Ludwig Heigenhauser is funded by the SFB 386 of the German Research Foundation.

Articles:

1. Paty DW, Ebers GC. Multiple Sclerosis. F.A. Davis Co., Philadelphia, 1998.

2. Weinshenker BG, Bass B, Rice GPA, Noseworthy J, Carriere W, Baskerville J, Ebers GC. The natural history of multiple sclerosis: geographically based study. I. Clinical course and disability. *Brain* 1989; **112**:133-46.

3. Miller DH, Khan OA, Sheremata WA, et al. A controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 2003; **348**:1598-99.

4. Rice GPA, Ebers GC. Interferons in the treatment of multiple sclerosis: do they prevent the progression of the disease? *Arch Neurol* 1998; **55**: 1578-80.

5. Ebers GC. Preventing multiple sclerosis? *Lancet* 2001; **357:** 1547.

6. Johnson KP, Brooks BR, Ford CC, et al. Sustained clinical benefits of glatiramer acetate in relapsing multiple sclerosis patients observed for 6 years. Copolymer 1 Multiple Sclerosis Study Group. *Mult Scler* 2000; **6**: 255-66.

7. Randomised double-blind placebo-controlled study of interferonbeta1a in relapsing/remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. *Lancet* 1998; **352**: 1498-504.

8. Goodkin DE. Inter and intra observer variability for grades 1.0-3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). *Neurology* 1992; **42**: 859-63.

9. Kappos L. Effect of drugs in secondary disease progression in patients with multiple sclerosis. *Mult Scler* 2004; **10**: S46-55.

10. SPECTRIMS Study. Randomized trial of interferon beta-1a in secondary progressive multiple sclerosis. 1. Clinical Results. *Neurology* 2001; **56**:1496-504.

11. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. European Study Group on interferon β-1b in secondary progressive MS. *Lancet* 1998; **352**: 1491-97.

12. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomised, double-blind, placebo-controlled trial. The IFNB Multiple Sclerosis Study Group. *Neurology* 1993; **43**: 655-61.

13. 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 multiples sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group. *Neurology* 1995; **45**: 1268-76.

14. Jacobs LD, Cookfair DL, Rudick RA, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis

Collaborative Research Group (MSCRG) *Ann Neurol* 1996; **39**: 285-94. Erratum in: *Ann Neurol* 1996; **40**: 480.

15. Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; **13**: 1444-52.

16. Noseworthy JH, Ebers GC, Vandevoort MK, Farquhar RE, Yetsir R, Roberts R. The impact of blinding on a randomised double blind placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994; 44:16-20

17. Cottrell DA, Rice GPA, Hader W, Baskerville J, Koopman WJ, Ebers GC. The natural history of multiple sclerosis: a geographically based study: 5. The clinical features and natural history of primary progressive multiple sclerosis. *Brain* 1999; **122**: 625-39.

18 Liu C, Blumhardt LD. Disability outcome measures in therapeutic trails of relapsing-remitting multiple sclerosis: effects of heterogeneity of disease course in placebo cohorts. *J Neurol Neurosurg Psychiatry* 2000; **68**: 450-57.

19. Rudick RA,Goodkin DE, Jacobs LD, et al. Impact of interferon beta-1a on neurologic disability in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). *Neurology* 1997; **49**: 358-63.

20. Francis DA, Bain P, Swan AV, Hughes RA. An assessment of disability rating scales used in multiple sclerosis. *Arch Neurol* 1991; **48**: 299-301.

21. Noseworthy JH, Vandevoort MK, Ebers GC. Interrater variability with the expanded disability status scale (EDSS) and functional systems(FS) in a multiple sclerosis clinical trial. *Neurology* 1990; **40**: 971-75.

22. Fillipini G, Munari L, Incorvaia B, Ebers GC, Polman C, D'Amico R, Rice GPA. Interferons in relapsing remitting multiple sclerosis: a systematic review. *Lancet* 2003; **361**: 545-52.

23. Noseworthy JH, Vandervoort MK, Hopkins M, Ebers GC. A referendum on clinical trial research in multiple sclerosis: The opinion of the participants at the Jekyll Island Conference. *Neurology* 1989; **39**: 977-81.

24. Rudick R, Antel J, Confavreux C, et al. Recommendations from the National Multiple Sclerosis Society Outcomes Assessment Task Force. *Ann Neurol* 1997; **42**: 379-82.

25. Sheridan C. Fast track to MS drug. *Nat. Biotechnol*. 2004; **22**:939-941.

Table 1- Clinical characteristics of the study populations

Clinical features	Data features	Subset 1		Subset 2		Subset 3	
MS type		RR	SP	RR	SP	RR	SP
	N	254	171	262	254	216	237
gender	male	70	72	64	111	54	106
	female	184	99	198	143	162	131
	male/N	0.28	0.42	0.24	0.44	0.25	0.45
Attacks last 2 years	NA	27	2	52	6	49	6
	median	3	1	3	1	3	1
	range	0-8	0-8	0-8	0-8	0-8	0-8
	mean	3.0	1.3	3.0	1.2	3.1	1.2
	sd	1.37	1.49	1.30	1.43	1.34	1.44
Duration in years	median	5.3	11.3	4.8	12.1	4.4	12.4
	range	0.7-34.8	2.0-37.3	0.7-37.7	1.3-37.3	0.7-37.7	1.8-37.3
	mean	7.0	12.8	6.9	13.5	6.6	13.7
	sd	5.76	7.37	6.13	7.77	5.97	7.75
Age at onset	median	28	30	29	29	29.5	28.5
	range	13-45	4-57	10-48	4-48	10-48	4-48
	mean	28.0	30.4	28.9	29.1	29.2	29.0
	sd	6.89	8.65	7.20	7.82	7.26	7.82
Age at study entry	median	35	43	36	43	36	43
	range	17-52	23-66	17-55	23-65	17-55	23-65
	mean	34.8	43.0	35.7	42.4	35.6	42.5
	sd	7.42	8.07	7.70	7.81	7.67	7.83
EDSS at entry	median	2.5	4.0	2.5	5.5	2.5	5.5
	range	0.0-5.5	1.5-5.5	0.0-6.5	3.0-6.5	0.0-6.5	3.0-6.5
	mean	2.6	4.4	2.7	5.3	2.6	5.3
	sd	1.20	0.79	1.34	1.06	1.28	1.05
Time to last obs. in months	median	24	33				
	range	15-51	15-39				
	mean	23.8	28.7				
	sd	6.64	8.69				

Subset 1 - at least 15 months every three months an observation and EDSS at baseline <6.0
Subset 2 - at least one observation at two years ± one month but no other restriction for values after baseline, missing values allowed.
Subset 3 - at least one observation at two years w/o potential relapses
obs. = observation and recording of EDSS score

Table 2: Worsening (progression or treatment failure-TF) and "improvement" (TI) events and sustainability in RR and SP MS for definitions of disability treatment failure /improvement used in MS. Analyses include all individuals in subset 1 without truncation

RRMS

Confim. Period	EDSS rise	progression (TF)		improvement (TI)		TI events / TF+TI
		No. events	not sust.*	No. events	not sust.*	
None	0.5	178	119	160	120	0.473
None	1.0	124	67	100	62	0.446
3 months	0.5	113	46	122	73	0.519
3 months	1.0	68	16	60	25	0.469
6 months	0.5	84	23	86	40	0.506
6 months	1.0	46	6	45	12	0.495
SPMS						
None	0.5	129	58	88	68	0.406
None	1.0	94	23	49	29	0.343
3 months	0.5	110	29	63	43	0.364
3 months	1.0	78	9	28	13	0.264
6 months	0.5	91	16	47	26	0.341
6 months	1.0	56	2	21	9	0.273

*Sustained (sust.) - increase or decrease in score did not remit by the final evaluation

Table 3: EDSS changes (N trial participants) for ½ point intervals from -2.5 - 5.0+ (in bold) over one/two years for subsets 1 and 3 in SP and RR trials*

subset	N =	**-2.5**	**-2.0**	**-1.5**	**-1.0**	**-0.5**	**0.0**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**	p-value
A	254	1	4	13	32	44	81	39	19	10	6	3	1	0	1	0	0	0.793
B	161	0	6	9	15	30	39	24	18	8	3	4	3	0	1	1	0	0.187
C	216	4	9	13	17	34	52	33	18	13	6	8	1	0	4	3	1	0.082
D	171	1	1	7	14	24	47	32	26	7	6	3	2	1	0	0	0	0.002
E	112	0	0	1	9	10	20	26	19	13	5	5	3	0	1	0	0	<0.001
F	237	1	0	1	14	18	70	58	34	15	13	8	4	0	1	0	0	<0.001

p-values for one-sided Wilcoxon
A: 1 year RR subset 1 (median: 0.0, mean: -0.016)
B: 2 year RR subset 1 (median: 0.0, mean: 0.124)
C: 2 year RR subset 3 (median: 0.0, mean: 0.194)
D: 1 year SP subset 1 (median: 0.0, mean: 0.237)
E: 2 year SP subset 1(median: 0.5, mean: 0.638)
F: 2 year SP subset 3 (median: 0.5 mean: 0.508)

Table 4: Statistical comparisons of survival curves for improvement vs. worsening as defined by the use of confirmation and degree of EDSS change (columns I and II) for the following patient groups: RR (all from subset 1), RR 2 years (data truncated at 2 yrs) RR w/o 1st (first data point omitted) and for SP (all from subset 1)

Confirmation	Rise	RR patients Subset 1	RR truncated at 2years	RR w/o 1st observation	SP (all)
none	0.5	0.920	0.632	0.356	<0.001
none	1.0	0.107	0.082	0.225	<0.001
3 months	0.5	0.131	0.148	0.503	<0.001
3 months	1.0	0.860	0.913	0.417	<0.001
6 months	0.5	0.337	0.372	0.424	<0.001
6 months	1.0	0.788	0.812	0.893	<0.001

Log rank p-values for differences between survival plots.

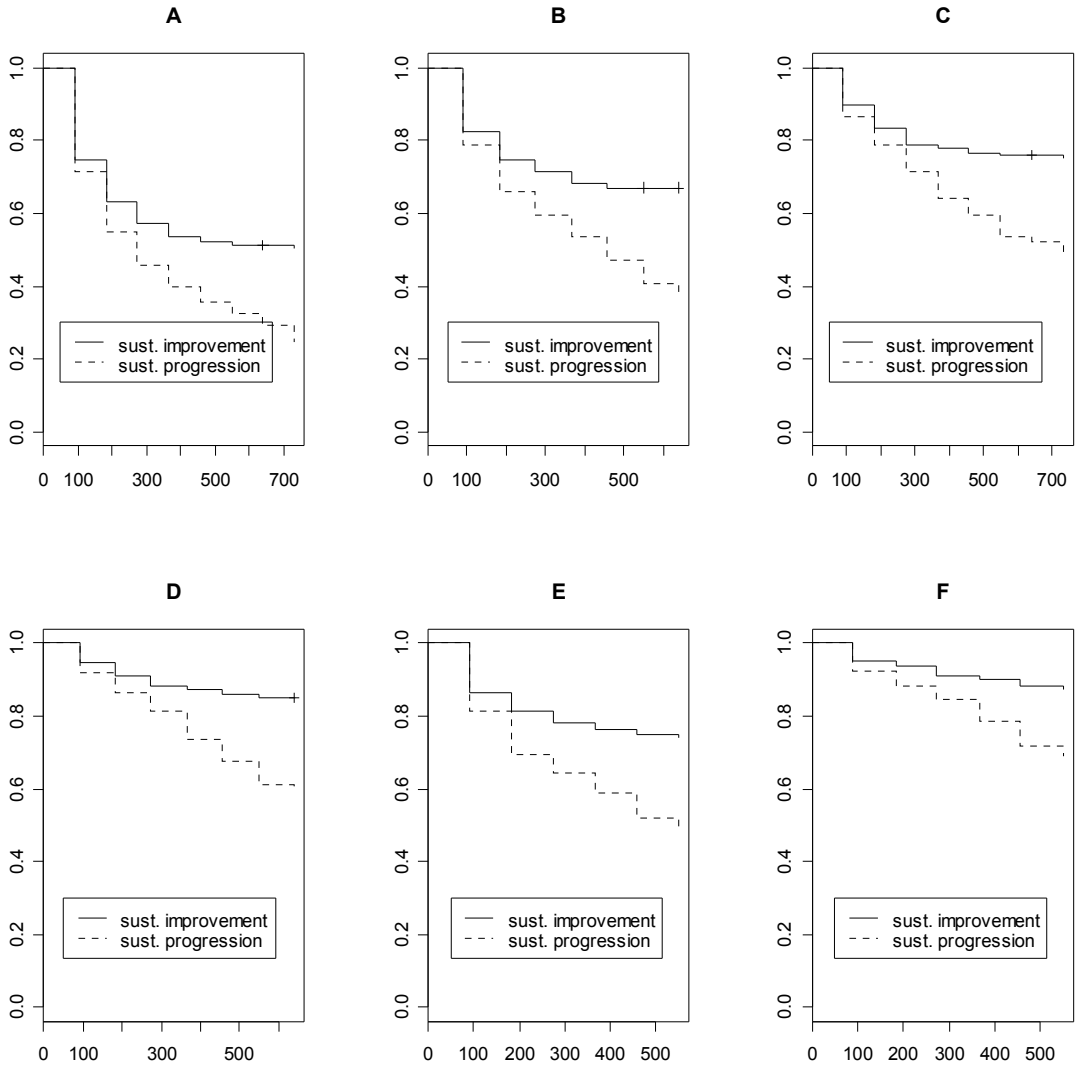

Figure 1: A-F Graphs for 171 subset 1 SP patients truncated after 2 years for the following outcomes by EDSS change and presence/timing of confirmation. X- axis in days, Y-axes - probability of not progressing/improving by same degree

A: no confirmation period, half point
B: 3 months confirmation period, half point
C: 6 months confirmation period, half point
D: no confirmation period, full point
E: 3 months confirmation period, full point
F: 6 months confirmation period, full point

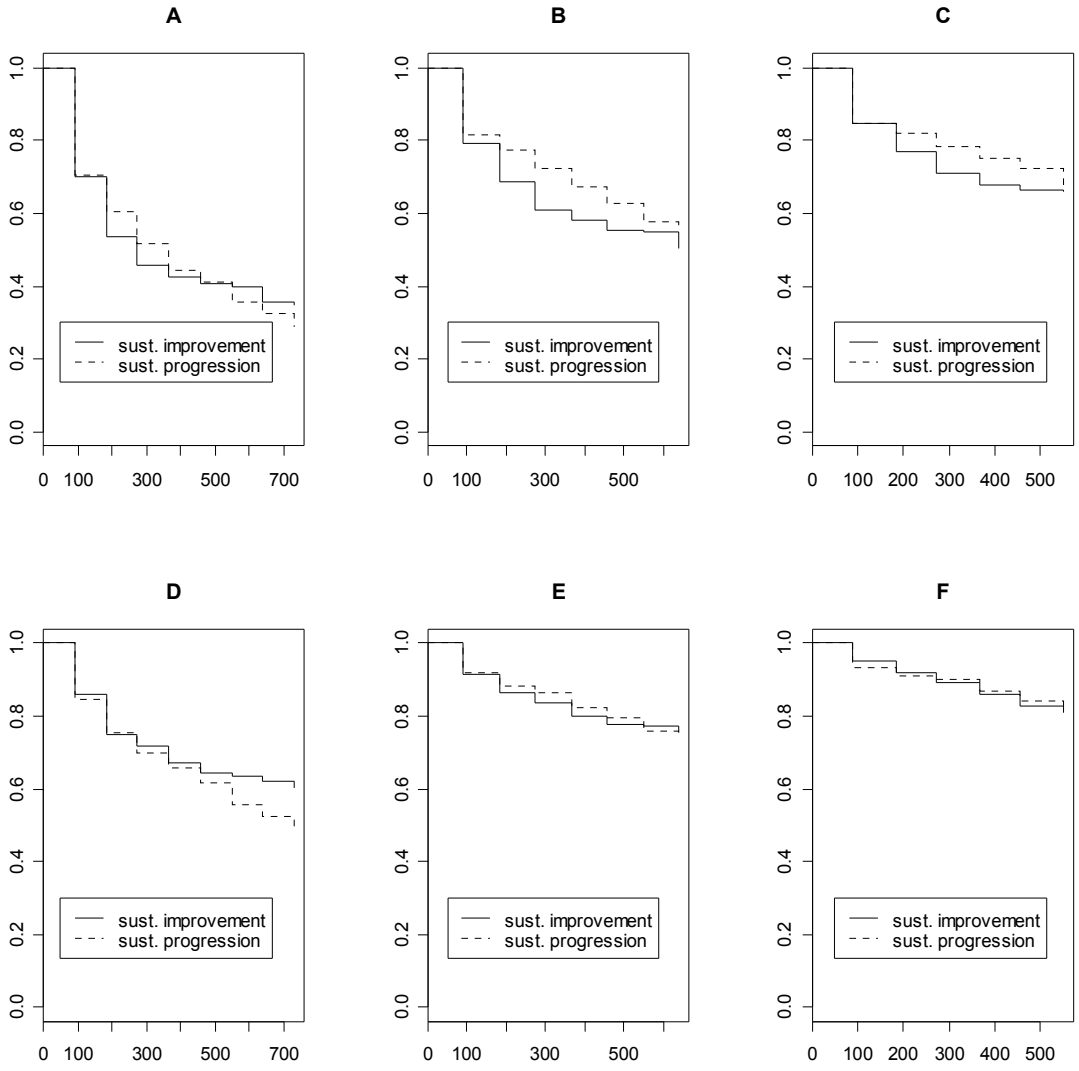

Figure 2: A-F Graphs for 254 RR patients subset 1; all observations were truncated after 2 years. X-axes in days. Y-axes - probability of not progressing/improving by same degree

A: no confirmation period, half point
B: 3 months confirmation period, half point
C: 6 months confirmation period, half point
D: no confirmation period, full point
E: 3 months confirmation period, full point
F: 6 months confirmation period, full point