Riluzole treatment, survival and diagnostic criteria in Parkinson plus disorders: The NNIPPS Study

Gilbert Bensimon,1 Albert Ludolph,2 Yves Agid,3 Marie Vidailhet,3 Christine Payan,1 P. Nigel Leigh4 and for the NNIPPS Study Group*

1 Département de Pharmacologie Clinique, Hôpital de la Pitié-Salpêtrière, Assistance Publique Hôpitaux de Paris & UPMC University Paris 06, UMR 7087, Paris, France
2 Department of Neurology, University of Ulm, Oberer Eselsberg, Ulm, Germany
3 Fédération de Neurologie, Hôpital de la Pitié-Salpêtrière, Assistance Publique Hôpitaux de Paris & UPMC University Paris 06, Paris, France
4 MRC Centre for Neurodegeneration Research, King’s College London, Institute of Psychiatry, PO 41, Department of Clinical Neuroscience, London, UK

*See appendix for the details on NNIPPS Study Group

Correspondence to: Prof. P. Nigel Leigh,
MRC Centre for Neurodegeneration Research,
King’s College London, PO 41, Institute of Psychiatry,
Camberwell, London SE5 8AF, UK
E-mail: n.leigh@iop.kcl.ac.uk

Parkinson plus diseases, comprising mainly progressive supranuclear palsy (PSP) and multiple system atrophy (MSA) are rare neurodegenerative conditions. We designed a double-blind randomized placebo-controlled trial of riluzole as a potential disease-modifying agent in Parkinson plus disorders (NNIPPS: Neuroprotection and Natural History in Parkinson Plus Syndromes). We analysed the accuracy of our clinical diagnostic criteria, and studied prognostic factors for survival. Patients with an akinetic-rigid syndrome diagnosed as having PSP or MSA according to modified consensus diagnostic criteria were considered for inclusion. The psychometric validity (convergent and predictive) of the NNIPPS diagnostic criteria were tested prospectively by clinical and pathological assessments. The study was powered to detect a 40% decrease in relative risk of death within PSP or MSA strata. Patients were randomized to riluzole or matched placebo daily and followed up to 36 months. The primary endpoint was survival. Secondary efficacy outcomes were rates of disease progression assessed by functional measures. A total of 767 patients were randomized and 760 qualified for the Intent to Treat (ITT) analysis, stratified at entry as PSP (362 patients) or MSA (398 patients). Median follow-up was 1095 days (range 249–1095). During the study, 342 patients died and 112 brains were examined for pathology. NNIPPS diagnostic criteria showed for both PSP and MSA excellent convergent validity with the investigators’ assessment of diagnostic probability (point-biserial correlation: MSA rpb = 0.93, P < 0.0001; PSP, rpb = 0.95, P < 0.0001), and excellent predictive validity against histopathology [sensitivity and specificity (95% CI) for PSP 0.95 (0.88–0.98) and 0.84 (0.77–0.87); and for MSA 0.96 (0.88–0.99) and 0.91 (0.86–0.93)]. There was no evidence of a drug effect on survival in the PSP or MSA strata (3 year Kaplan–Meier estimates PSP-riluzole: 0.51, PSP-placebo: 0.50; MSA-riluzole: 0.53, MSA-placebo: 0.58; P = 0.66 and P = 0.48 by the log-rank test, respectively), or in the population as a whole (P = 0.42, by the stratified-log-rank test). Likewise, rate of progression was similar in both treatment groups. There were no unexpected adverse effects of riluzole, and no significant safety concerns. Riluzole did not have a significant effect on survival or rate of functional deterioration in PSP or MSA, although the study reached over 80% power to detect the hypothesized drug effect within strata. The NNIPPS diagnostic criteria were consistent and valid. They can be used to distinguish between PSP and MSA with high accuracy, and should facilitate research into these conditions relatively early in their evolution.
Introduction

Progressive supranuclear palsy (PSP) and multiple system atrophy (MSA) are disabling and fatal neurodegenerative disorders for which no disease-modifying treatment is available. For the majority of patients with PSP and MSA, many of whom present with an atypical Parkinsonian or akinetic-rigid syndrome (‘Parkinson plus disorder’) the course is one of relentless progression, increasing disability and death with a median survival of 5–10 years from onset of symptoms (Litvan et al., 1996a; Testa et al., 1996, 2001; Ben-Shlomo et al., 1997; Schrag et al., 1999, 2008; Litvan, 2003; Golbe and Ohman-Strickland, 2007).

PSP and MSA have similar prevalence rates estimated at 2–7 per 100,000 person years (Golbe et al., 1988; Ben-Shlomo et al., 1997; Bowler et al., 1997; Schrag et al., 1999; Nath and Burn, 2000; Nath et al., 2001; Vanacore et al., 2001a, b; Watanabe et al., 2002). These are probably underestimates, however, because current diagnostic criteria are based on retrospective clinicopathological studies (Litvan et al., 1996a, b, d, 2003; Litvan, 2003) and both delayed diagnosis and mis-diagnosis are common. Although published consensus diagnostic criteria are highly specific they are relatively insensitive and a definite diagnosis can only be made through histopathology (Litvan et al., 2003). Since it is likely that neuroprotective strategies are best tested at a relatively early stage of disease, more sensitive diagnostic criteria are required for trials of potential disease-modifying agents. Although PSP and MSA often present as akinetic-rigid syndromes, each has distinctive pathological features. In MSA, a key feature is glial cytoplasmic inclusions with accumulation of α-synuclein in oligodendrocytes and neurons (Papp et al., 1989; Lantos and Papp, 1994; Spillantini and Goedert, 2000) whilst in PSP the hallmark is accumulation of abnormally phosphorylated microtubule-associated tau protein (τ) in neurons and glia (Hauw et al., 1994; Dickson et al., 2007). Although the pathogenic mechanisms underlying MSA and PSP are unknown, there is evidence that glutamate toxicity may contribute to neuronal damage in these and other neurodegenerative diseases (Albin and Greenamyre, 1992; Albers and Augood, 2001; Mattson, 2003; Przedborski, 2005). The benzothiazole drug riluzole has a number of pharmacological effects that contribute to neuroprotection in experimental paradigms of neurodegenerative diseases including anti-excitotoxic activity, blocking of voltage dependent sodium channels, free-radical scavenging, anti-apoptotic and neurotrophic effects and inhibition of protein aggregation (Doble, 1999; Heiser et al., 2002; Yoo et al., 2005; Caumont et al., 2006; Shortland et al., 2006). In a rodent model of MSA, riluzole improved some measures of neuronal damage (Diguet et al., 2005; Scherfler et al., 2005). Riluzole (up to 200 mg daily) is well tolerated and prolongs survival in amyotrophic lateral sclerosis (Bensimon et al., 1994; Lacomblez et al., 1996; Miller et al., 2007). Thus far, riluzole remains the only agent shown to modify disease progression in a human neurodegenerative disorder.

In order to test the hypothesis that riluzole may slow disease progression in PSP and MSA, we carried out a phase-III, randomized double-blind placebo controlled trial in 44 centres in France, Germany and the UK. The study design incorporated ancillary objectives including natural history, development and validation of more sensitive diagnostic criteria and functional measures of disease severity and progression, cognition, quality of life, health economics, MRI changes, pathology and the establishment of brain and DNA banks. We describe here the design and main outcomes of the NNIPPS trial in terms of the efficacy and safety of riluzole, the psychometric validity of the NNIPPS diagnostic criteria in relation to clinic and pathology and the major factors influencing prognosis for PSP and MSA.

Methods

NNIPPS was designed as a double-blind placebo-controlled, stratified (by diagnosis of MSA or PSP, and by centre), parallel group, European (France, Germany, United Kingdom) trial assessing the efficacy and safety of riluzole at flexible dose (50–200 mg/day) (Fig. 1). The primary objective of NNIPPS was to demonstrate the efficacy of riluzole on survival (primary end-point) and rate of decline in motor function (secondary end-points).

The number of patients required was determined for the primary end-point survival and for each stratum (PSP, MSA). Assumptions included mean disease duration of 3 years prior to entry and no loss to follow-up. Assuming a 41% death rate at 3 years in the placebo group (Litvan et al., 1996a; Testa et al., 1996, 2001; Ben-Shlomo et al., 1997; Litvan, 2003; Golbe and Ohman-Strickland, 2007; Schrag et al., 2007), 400 patients provide over 80% power to detect a 40% decrease in the relative risk of death in the treated group using the log-rank test with two-sided α risk set at 0.05. With both strata combined (800 patients), the power reached 98% to detect a 40% decrease in the risk of death with active treatment compared to placebo, assuming that a minimum of 272 events would be observed over the 36-month trial period.

Patients and treatment

From consensus criteria (Litvan et al., 1996a, d, 2003) we derived simplified operational diagnostic criteria suitable for large-scale clinical
trials, with the aim of maximizing sensitivity for recruitment. Strata were defined according to these operational diagnostic criteria (Table 1). The history of the condition in each patient was evaluated at entry using systematic questionnaires recording the initial (pre-symptom) symptom, and current syndromes (entry and last visit). Response to levodopa therapy was evaluated at entry. Additional assessments related to ancillary studies including a new functional scale (the Parkinson Plus scale), magnetic resonance imaging, neuropsychology, health economics and quality of life, genetic testing and neuropathology, will be presented separately.

Patients were allocated to treatment according to a computer generated randomization list, stratified for diseases (MSA or PSP) and clinical centre, with a riluzole to placebo ratio of 1:1. Riluzole (50 mg) or placebo was prepared as identical tablets (Sanofi-Aventis, Antony-France). Packaging and labelling (LC², Lentilly-France) and treatment management (Cardinale, Corby-UK; Clindata, Weilerswist-Germany; AGEPS-AP HP, Paris-France) was performed so as to safeguard blinding to treatment allocation throughout the trial duration.

Following randomization, a monthly dose-titration over 3 months was used with increasing dosage from one, two and four tablets daily during which time tolerance was assessed with monthly laboratory tests for haematology (full blood count) and liver function (including ALT and AST) and patient’s reporting of adverse events. Dose flexibility (one to four tablets per day) according to tolerance was allowed throughout the 36-month study period, with each dose-adjustment recorded. Study treatment withdrawal was not considered as an ‘end of study’ and the protocol required that patients should be followed for the ITT analysis to the end of the planned double-blind period. Treatments were delivered to patients every 3 months and the tablets returned were counted to evaluate compliance.

Assessments
To assess the psychometric validity (see Statistical analysis section) of the NNIPPS operational diagnostic criteria, investigators were required to assign at entry a diagnostic probability (PSP or MSA) for each patient using two 100 mm visual analogue scales (VAS). These clinical diagnostic assessments were completed at entry, every 12 months thereafter, and at the last visit. When possible, clinical diagnosis was compared to neuropathological diagnosis, which was assessed blind to the clinical diagnosis at all stages of processing and analysis of donated brains. The latter were processed according to a standard protocol incorporating formalin fixation of one hemisphere (randomly allocated by the UK, French or German coordinating centres) with freezing (at ~80°C) of the other hemisphere for banking. Tissue sections were assessed against standard diagnostic criteria (Lantos and Papp, 1994; Litvan et al., 1996d; Dickson, 1999) with cross-examination and consensus scoring for each case.

The primary criterion of efficacy was defined as survival during the 36-month double-blind period of the study (1095 days included) or until the administrative cut-off date (November 30, 2004 included) whichever came first. The primary end-point was defined as death from any cause. All dates of death were documented with death certificates. For all surviving patients, the date of last contact was also documented.

Secondary end-points included standard assessments used in idiopathic Parkinson’s disease completed at entry and 6 monthly, using the Hoehn and Yahr staging scale (Hoehn and Yahr, 1967), the Schwab and England Activities of Daily Living scale (SEADL, Schwab and England, 1969) as well as generic health scale assessments using the Clinical Global Impression for disease severity completed by investigators (CGI-ds; Streiner and Norman, 2003) and a CGI adapted to autonomic function assessment (CGI-dysautonomia). The Mini-Mental State Examination (MMSE, Folstein et al., 1975) was completed by the clinical observer at entry and every 12 months. In addition, a specific functional measure assessing ambulation was developed (Short Motor Disability Scale, SMDS) and performed at entry and 3 monthly (Supplementary Table 1). Anticipated non-serious adverse events of riluzole included dizziness, gastrointestinal symptoms, and fatigue (Lacomblez et al., 1996). The main anticipated serious adverse event (SAE) for riluzole was serious abnormality of liver function tests defined as ALT >5× the upper limit of normal. Safety was evaluated through clinical examination, vital signs, routine laboratory tests including haematology and transaminases (AST, ALT), concomitant medication and patient reports of adverse events (at entry and 3 monthly). Adverse events, serious and non-serious were coded using MedDRA® version 6.0 (MedDRA MSSO, Northrop Grumman Corp., Reston, VA, USA). Weight and electrocardiogram were assessed at entry and at the final visit.

Statistical analysis
An Independent Data Monitoring and Safety Committee (IDMSC) was established for unblinded review of all SAEs during the trial and to advise the Steering Committee at regular intervals regarding continuation of the trial according to predefined stopping rules. No interim analysis for efficacy was planned but four safety analyses were performed throughout the study to ensure that mortality in the treated group was not in excess compared to the placebo group.

The detailed statistical plan was submitted to the French IRB prior to unblinding. The primary analysis was conducted following the ITT principles. The ITT population was defined as all randomized subjects who received at least one dose of study medication and for whom there were no major violations of GCP (ICH-E6). Sensitivity analyses population definition is described in Supplementary Text 1.

For the primary analysis, the diagnosis at inclusion was used to define the PSP and MSA strata. In addition, the statistical plan defined...
sensitivity analyses by diagnosis at the end of the study to allow for the possibility of misdiagnosis.

The Safety Population comprised all randomized subjects who received at least one dose of study medication. All analyses were performed using SAS Software version 11. According to the guidelines for the standard for educational and psychological testing (American Psychological Association, 1985) two facets of the psychometric validity of the diagnostic criteria were evaluated, the convergent and the predictive validity. The convergent validity was assessed by the degree of correlation between diagnostic classification according to inclusion criteria and the investigator’s assessment of diagnostic probability on the VAS. As inclusion criteria represent a nominal variable with two modalities (MSA, PSP), we used the point biserial coefficient which is the relevant method in this case (point-biserial correlation, \( r_{pb} \); Nunnally and Bernstein, 1994). Predictive criterion-related validity of the clinical diagnostic criteria was assessed by calculating the specificity, sensitivity, percent correct classification and positive likelihood (Attia, 2003), using pathological diagnosis as the gold standard.

Descriptive analyses summarized the overall population, and the population sub-divided by treatment group, by strata (PSP versus MSA) and by treatment within strata. Categorical data were summarized by frequency and percentage. The log-linear model was used to compare distributions between factors including treatment, strata, country and all interactions (chi-square tests of partial association) (Bishop et al., 1975). Continuous data were summarized by mean and standard deviation. Between treatment groups comparison at entry was carried out using three-factor variance analyses including treatment, strata, country and all interaction factors (Winer, 1971).
For the primary end-point survival, between groups survival curves were compared using the Mantel-Cox (log-rank) test (Mantel and Haenszel, 1958). Treatment effect was assessed with the stratified log-rank test ($p_{\text{two-tailed test}} < 0.05$). The Cox model (Cox, 1972) including treatment, strata and interaction factors, was used to check for treatment by strata (MSA, PSP) interaction.

The influence of demographic and clinical variables at entry on survival was tested with univariate and multivariate Cox proportional-hazard analysis. Multivariate analysis used an automatic up and down stepwise selection of variables (Allison, 1995).

Secondary end-point variables used for assessing disease progression included the SMDS, the SEADL, the Hoehn and Yahr staging and the CGI-ds. For each patient, repeated measurements were summarized by slope of change over time using linear regression methods (unweighted least square estimate; Wu, 1988). Comparisons of slopes between the treatment groups were performed using three-factor variance analyses, including strata, country, treatment and interaction factors. Serious and non-serious adverse events were compared between treatment groups with the Pearson’s chi-square test or Fisher exact test where appropriate.

**Ethics**

Prior to inclusion, patients gave their informed written consent to participate in the study. Separate consent was obtained both for DNA sampling and for post-mortem brain tissue donation. The protocol and subsequent amendments were approved by Ethics Committees/Institutional Review Boards of each coordinating centre in the three participating countries. The trial was conducted according to International standards of Good Clinical Practice-(ICH guidelines and the Helsinki Declaration). An internal audit of the study was carried out at the end of the trial before unblinding by an independent auditor (Qualilab, Olivet, France) in the nine largest centres (three in France, three in Germany and three in UK) accounting for 43% of the overall trial population, and the primary criteria survival was audited and appropriate documentation certified as 100% complete over the whole study population.

**Results**

During the study period, April 2000 to December 2004, the Independent Data Monitoring and Safety Committee performed four safety analyses (last review, June 2004) with advice to continue the trial on each occasion.

**Study populations**

From April 2000 to July 2002, a total of 767 patients (PSP, $n = 363$; MSA, $n = 404$) were recruited in 44 centres with 19 centres in France ($n = 320$), 12 in Germany ($n = 243$) and 13 in the UK ($n = 204$). Two patients never received treatment, and therefore were excluded from all analyses. For five additional patients, signed informed consent ($n = 3$) or code break envelopes ($n = 2$) were missing and therefore these patients were excluded from the ITT population (All, $n = 760$; PSP, $n = 362$; MSA, $n = 398$) but remained in the safety population analysis ($n = 765$) (Fig. 2). The mean daily dose on treatment in ITT population was 3.6 tablets (180 mg riluzole or placebo), and overall mean compliance was

![Fig. 2 NNIIPPS populations in analyses.](image-url)
Table 2 ITT population characteristics at entry

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Female (%)</td>
<td></td>
<td>Female (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40.3</td>
<td>44.2</td>
<td>44.8</td>
<td>46.2</td>
</tr>
<tr>
<td>Age at entry (years)</td>
<td>67.7 ± 70</td>
<td>62.6 ± 8.1</td>
<td>68.0 ± 6.6</td>
<td>61.9 ± 8.5</td>
</tr>
<tr>
<td>Age at onset (years)</td>
<td>63.8 ± 70</td>
<td>58.2 ± 8.3</td>
<td>63.9 ± 70</td>
<td>57.3 ± 8.6</td>
</tr>
<tr>
<td>Disease duration (years)</td>
<td>3.9 ± 1.9</td>
<td>4.4 ± 2.2</td>
<td>4.1 ± 1.9</td>
<td>4.5 ± 1.9</td>
</tr>
<tr>
<td>Short Motor Disability</td>
<td>6.4 ± 3.7</td>
<td>6.1 ± 3.9</td>
<td>6.7 ± 3.6</td>
<td>6.1 ± 3.6</td>
</tr>
<tr>
<td>Assessment Battery</td>
<td>11.1 ± 4.3</td>
<td>14.6 ± 3.2</td>
<td>11.3 ± 4.1</td>
<td>14.3 ± 3.3</td>
</tr>
<tr>
<td>Mini-Mental Status</td>
<td>25.4 ± 4.4</td>
<td>27.8 ± 2.3</td>
<td>25.2 ± 4.4</td>
<td>27.6 ± 2.5</td>
</tr>
<tr>
<td>Schwab &amp; England Activity</td>
<td>50.2 ± 24.5</td>
<td>53.8 ± 24.8</td>
<td>48.3 ± 23.6</td>
<td>53.3 ± 24.3</td>
</tr>
<tr>
<td>Hoehn and Yahr Staging</td>
<td>3.6 ± 1.0</td>
<td>3.4 ± 1.0</td>
<td>3.6 ± 0.9</td>
<td>3.5 ± 1.0</td>
</tr>
<tr>
<td>Clinician Global Impression Disease severity</td>
<td>3.6 ± 1.0</td>
<td>3.6 ± 1.0</td>
<td>3.7 ± 1.0</td>
<td>3.6 ± 0.9</td>
</tr>
<tr>
<td>Clinician Global Impression</td>
<td>0.6 ± 0.6</td>
<td>1.8 ± 0.8</td>
<td>0.6 ± 0.6</td>
<td>1.8 ± 0.8</td>
</tr>
</tbody>
</table>

Quantitative variables were analysed using variance analysis and categorical data using a log-linear model. Factors included in models were strata, treatment, country and strata by treatment interactions. Significance was set at *P* < 0.05 (two-tailed test) for each factor or interaction. No differences related to treatment group were detected at entry for the whole population or within strata.

All values are mean ± SD.

a *n* = 180 due to one missing value.

* Differences between PSP and MSA strata (*P* < 0.05; two-tailed test) were found for all variables except for sex and the Short Motor Disability Scale score.

81.2% ± 31.5. In the riluzole group, the maximum tolerated dose of riluzole for the 284 (75%) patients who did not stop study treatment until death or study cut-off was 200 mg in 237 patients (83.4%), 150 mg in 7 patients (2.5%), 100 mg in 25 patients (8.8%) and 50 mg in 15 patients (5.3%).

The sensitivity analysis population is described in Supplementary Text 1.

Effect of riluzole in patients with PSP and MSA

Treatments were well balanced in the overall population and within strata. There was no significant difference between treatment groups in terms of demographic features, disease severity, previous medical history or concomitant disease at entry in the overall population or within strata (Table 2).

Follow-up at cut-off date was complete and documented in all patients regardless of treatment compliance except for three patients, two of whom withdrew consent in order to be included in another trial and one who underwent medically assisted suicide. At the cut-off date, the surviving patients had a mean (±SD) follow-up time from randomization of 1055 ± 88 days with 81.1% having completed 3 years double-blind follow-up (1095 days). Overall 140 patients (placebo *n* = 71, riluzole *n* = 72) had less than 3 years follow-up at the administrative cut-off date, and three had to be censored at the time they withdrew from the study, as mentioned above. The mean time in study for these 143 patients (18.8%) was 979.4 ± 117.4 days (placebo: 970.5 ± 127.3 days; riluzole: 988.2 ± 106.9 days).

Overall 342 patients (45.0%) died during the double-blind period, with no difference between the PSP and the MSA strata [PSP, *n* = 171 (47.2%); MSA, *n* = 171 (43.0%); *P* = 0.21 by the log-rank test]. With the strata combined, Kaplan–Meier survival estimates at 36 months for riluzole and placebo groups were 52.6% and 54.9%, respectively. Comparison of survival curves in treatment groups showed no statistically significant difference either in the overall population (*P* = 0.42 by the stratified log-rank test) or within strata (Kaplan–Meier estimates for PSP–riluzole: 0.51, PSP-placebo: 0.50; and for MSA-riluzole: 0.53, MSA-placebo: 0.58; *P* = 0.66 and *P* = 0.48 by the log-rank test, respectively) (Fig. 3). Accordingly there was no statistically significant treatment by strata interaction (*P* = 0.85, by a Cox model analysis).

For secondary efficacy end-points of disease progression, 651 patients (86%) had at least two assessments allowing calculation of a slope of change in scores for Hoehn and Yahr Staging, SEADL and CGI for severity. For SMDS 714 (94%) patients had usable data for calculating slope of change. All scales showed high sensitivity to change with time (*P* < 0.0001). Strata differences were evident with the SEADL, CGI for severity and SMDS (*P* = 0.003, *P* = 0.0003, *P* = 0.03, respectively; Table 3), with the PSP group showing more rapid progression compared to the MSA group.

As with the primary end-point, there was no statistically significant difference between treatment groups in progression rate with any of the scales, either in the overall population or within strata. The PPPT analysis showed identical results for primary and secondary end-points (data not shown).

To test our initial assumptions, using the data acquired during the study, we calculated the detectable difference in survival and functional change. With power greater than 0.8, and an α risk at 0.05 (two-tailed), the actual number of patients recruited and events observed would have allowed us to detect a 40% decrease in relative risk of death for PSP and 35% for MSA, consistent with our initial hypothesis. For functional change, using the SEADL as the most sensitive scale, we would have been able to detect differences in annual rate of progression of 32% for PSP and 36% for MSA.
Safety

Analysis of all SAEs revealed no significant differences between riluzole and placebo groups for the frequency of unexpected SAEs. However, for expected SAEs (other than death) with frequency of greater than 5%, gastrointestinal events were more common with riluzole treatment (Table 4). Borderline significant differences were observed in SAEs with frequency below 5%. Cardiac disorders were more common in the riluzole group (2.6% versus 0.8% placebo) and urinary disorders more frequent in the placebo group (4.4% versus 1.8% riluzole). With all treatment groups combined, there were significant differences between PSP and MSA, with injuries due to falls more common in PSP (18% versus 8% in MSA) and gastrointestinal disorders related to dysphagia more frequent in PSP (11% versus 6%). Events related to death are given in Table 4. The main causes were related to respiratory disorders, general condition (terminal state) and infections (urinary or pulmonary). Gastrointestinal complications were more common in PSP (4% versus 1% for MSA). Among the non-serious events, gastrointestinal disorders were again more common with riluzole (34% versus 27% with placebo, \( p = 0.03 \)). The frequency of vascular disorders (mainly haematoma related to falls) was also slightly higher in the riluzole group (13% versus 8% with placebo, \( p = 0.02 \)). Elevation of transaminases was rare, with values above three times normal reported for only 11 patients on riluzole (3%) and eight patients on placebo (2%).

Validity of NNIPPS diagnostic criteria

At entry, the investigators’ assessments of diagnostic probability for PSP or MSA were in close agreement with the inclusion category (point-biserial correlation for MSA \( r_{pb} = 0.93, P < 0.0001 \), and for PSP \( r_{pb} = 0.95, P < 0.0001 \)) demonstrating an excellent convergent validity. For a few patients, the VAS probability was equal

---

Table 3 Slope of change in functional measures

<table>
<thead>
<tr>
<th></th>
<th>Placebo, ( n = 181 )</th>
<th>MSA, ( n = 199 )</th>
<th>Riluzole, ( n = 181 )</th>
<th>MSA, ( n = 199 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short Motor Disability Scale</strong></td>
<td>3.1 ± 4.4 (71)</td>
<td>2.3 ± 2.7 (190)</td>
<td>3.0 ± 4.0 (68)</td>
<td>2.7 ± 2.9 (185)</td>
</tr>
<tr>
<td><strong>Schwab and England Activity Daily Living</strong></td>
<td>−16.3 ± 17.5 (162)</td>
<td>−11.5 ± 12.4 (172)</td>
<td>−15.2 ± 14.1 (149)</td>
<td>−12.9 ± 16.5 (68)</td>
</tr>
<tr>
<td><strong>Hoehn and Yahr Staging</strong></td>
<td>0.5 ± 0.8 (162)</td>
<td>0.5 ± 0.6 (172)</td>
<td>0.6 ± 0.7 (49)</td>
<td>0.5 ± 0.8 (168)</td>
</tr>
<tr>
<td><strong>Clinical Global Impression of disease severity</strong></td>
<td>0.7 ± 0.7 (162)</td>
<td>0.5 ± 0.6 (171)</td>
<td>0.7 ± 0.8 (48)</td>
<td>0.5 ± 0.8 (168)</td>
</tr>
</tbody>
</table>

Slope of change for each patient was calculated using simple linear regression method of dependent variable with time since randomization. Patient with at least two measures in study were included in analyses. Slope of change in functional scale scores are shown as mean points/year ± SD with numbers of patients in each analysis in parentheses. No statistically significant difference between treatment groups could be evidenced in either scale.

* Strata differences were observed with the SEADL (\( p = 0.003 \)), CGI disease severity (\( p = 0.0003 \)) and SMDS (\( p = 0.03 \)).
for PSP and MSA (Fig. 4). Overall, diagnostic confidence (VAS probability of each diagnosis) was significantly higher (Student’s t-test 2.89, P = 0.004) for PSP patients [mean 81.2, SD ± 12.8 (range 0.40–1)] than for MSA patients [mean 78.4, SD ± 14.1 (range 0.21–1)].

A total of 210 patients (27%) consented to brain donation. Analysis of 112 brains was completed at the time of writing. Comparisons of patients with a pathological diagnosis with those dying without pathological studies showed no difference in demographic characteristics, disease severity at entry or diagnostic probability (data not shown).

Histopathology showed that the NNIpps clinical diagnostic criteria had correctly identified Parkinson plus syndromes in 94% of cases. Two patients with pathologically confirmed MSA were mis-stratified at entry as PSP, and three with pathologically confirmed PSP had been mis-stratified as MSA (overall 4.9% mis-stratification). For the PSP stratum, seven cases were found to have other conditions, including two with Lewy body disease, one with amyotrophic lateral sclerosis (ALS), one with basophilic inclusion body disease, and three with corticobasal degeneration (CBD). In the MSA stratum, there were three misdiagnoses including one case with Lewy body disease, one with ALS and one with non-specific lesions. The sensitivity (95% CI) and specificity (95% CI) of the NNIpps clinical diagnostic criteria for MSA were 0.96 (0.88–0.99) and 0.91 (0.86–0.93) respectively, with a correct clinical diagnosis in 0.93 (0.87–0.96) of cases and positive likelihood of 10.67 (6.28–14.39) (Table 5). For patients diagnosed clinically as PSP, but excluding cases diagnosed pathologically as CBD, sensitivity (95% CI) and specificity (95% CI) of clinical diagnostic criteria were 0.95 (0.89–0.98) and 0.84 (0.77–0.87) respectively, with 0.89 (0.83–0.93) having a correct clinical diagnosis and positive likelihood of 5.79 (3.79–7.58). These results are evidence of an excellent predictive validity of the NNIpps diagnostic criteria. This was not explained by the higher than expected disease severity and thus the large number of late stage patients in the overall population. We conducted the same analysis on the population broken down by severity according to the median of the CGI-ds at baseline, which was identical in the overall population and in the sub-group with neuropathology diagnosis. Mean ± SD CGI-ds was 2.6 ± 0.5 and 2.9 ± 0.3 in early patients with PSP and MSA, respectively, and 4.5 ± 0.6 and 4.5 ± 0.7 in late patients with PSP and MSA, respectively. As shown in Table 5, results in early patients were not significantly different from those in late patients demonstrating a good consistency of the parameters with regard to disease progression. All three CBD cases fell within the late stage category. When CBD was included in the PSP neuropathology cases, there was slightly but not significantly increased diagnostic specificity, overall fraction correct, and positive likelihood for the overall sample and the late disease patients subgroup (data not shown).

Reassessment of the clinical diagnosis was achieved at least once during the trial period in 554 (73%) patients who did not have a neuropathological diagnosis (PSP, 71%; MSA, 75%). The overall rate of change in clinical diagnosis after entry was 7% (Table 6), consistent with the neuropathological findings.

Table 4 Serious Adverse Events—MedDRA classification (By System Organ Class)

<table>
<thead>
<tr>
<th>System Organ Class, n (%)</th>
<th>Related to death</th>
<th>Related to hospitalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placebo N = 383</td>
<td>Riluzole N = 382</td>
</tr>
<tr>
<td>Respiratory, thoracic and mediastinal disorders</td>
<td>60 (25)</td>
<td>65 (29)</td>
</tr>
<tr>
<td>General disorders and administration site conditions</td>
<td>56 (24)</td>
<td>56 (25)</td>
</tr>
<tr>
<td>Infections and infestations</td>
<td>37 (16)</td>
<td>49 (22)</td>
</tr>
<tr>
<td>Cardiac disorders</td>
<td>25 (11)</td>
<td>14 (6)</td>
</tr>
<tr>
<td>Surgical and medical procedures</td>
<td>13 (6)</td>
<td>8 (4)</td>
</tr>
<tr>
<td>Gastrointestinal disorders</td>
<td>13 (6)</td>
<td>7 (3)</td>
</tr>
<tr>
<td>Nervous system disorders</td>
<td>12 (5)</td>
<td>7 (3)</td>
</tr>
<tr>
<td>Metabolism and nutrition disorders</td>
<td>3 (1)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>Injury, poisoning and procedural complications</td>
<td>6 (3)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Neoplasms benign, malignant and unspecified</td>
<td>5 (2)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Psychiatric disorders</td>
<td>2 (1)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Skin and subcutaneous tissue disorders</td>
<td>3 (1)</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Vascular disorders</td>
<td>–</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Renal and urinary disorders</td>
<td>2 (1)</td>
<td>–</td>
</tr>
<tr>
<td>Reproductive system and breast disorders</td>
<td>1 (0.4)</td>
<td>–</td>
</tr>
<tr>
<td>Hepatobiliary disorders</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Investigations</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Musculoskeletal and connective tissue disorders</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Blood and lymphatic system disorders</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Eye disorders</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ear and labyrinth disorders</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total events</td>
<td>238</td>
<td>224</td>
</tr>
<tr>
<td>Total patients, Number of events (percentage of patients)</td>
<td>169 (44)</td>
<td>176 (46)</td>
</tr>
</tbody>
</table>

a Statistically significant difference between treatment by the Fischer exact test.
Clinical features and natural history of the NNIPPS cohort

Analysis of the systematic questionnaire on initial clinical signs showed that the akinetic-rigid syndrome (with or without falls) was a frequent presenting syndrome in the PSP stratum (70.2%) and in the MSA stratum (61.8%). In the PSP stratum, oculomotor abnormalities were an initial feature in only 7.7% of patients, while 11.0% had presented with a behavioural or cognitive syndrome, and 5.8% had bulbar or pseudo-bulbar features. In MSA, second to akinetic-rigid syndrome, the most common presenting clinical features were cerebellar (22.1%) and genito-urinary (9.1%). The date of onset of gait instability or falls was broken down by disease severity as defined by the CGI. Early patients were defined as those below the median (CGI, range 1–3); late patients were those equal or over the median (CGI, range 4–6).

The date of onset of gait instability or falls was documented in 93.3% of the population (92.5% PSP, 94.0% MSA). Falls within the first year of disease onset, as incorporated in the NINDS-SPSP criteria (Litvan et al., 2003) were present in only 49.6% of the PSP patients, and were also present in 21.9% of the MSA group. Similar results were observed in neuropathologically confirmed cases (PSP 53.6%, MSA 27.9%). A similar proportion of patients in the PSP and MSA strata had levodopa therapy at entry [PSP, n = 307 (85%); MSA, n = 334 (84%)], although the mean daily dose of levodopa was higher in the MSA group (636 mg/day, range 50–2100 mg) compared to the PSP group (509 mg/day, range 50–1600 mg) (P < 0.0001) (Supplementary Table 2). Overall, most patients had a very poor response to levodopa therapy. A greater than 50% response to levodopa was reported for only 1.5% of MSA patients, and none of the PSP group. A best-ever response to levodopa therapy >50% was reported more frequently for MSA patients (9.1%) than for PSP patients (2.6%) (P = 0.0002). In the majority of those who had experienced a good response to levodopa, the duration of response was <1 year.

At time of entry in the study, the distribution of the various syndromes is given for both diagnostic categories (Fig. 5). Oculomotor abnormalities were present in all PSP patients except one, but were also noted in 19% of MSA patients. PSP patients showed a higher rate of cognitive and behavioural syndromes. Dysautonomia was present in the majority of MSA patients, with urinary symptoms (87%) more common than cardiovascular symptoms (57%). Urinary symptoms were also present in 48% of PSP patients. A cerebellar syndrome was reported in 50% of MSA cases, but also a small number of PSP patients (6%). Pyramidal signs were present in approximately half the cases, slightly more in the MSA strata. A high frequency of bulbar/pseudobulbar features was reported in the MSA patients group (63%) and the PSP group (76%).

There were no differences between disease strata in terms of gender, weight or height. MSA patients were younger than the PSP patients at entry, younger at disease onset and had longer disease duration prior to entry (Table 2). Assessments of disease severity by the CGI-ds or the modified Hoehn and Yahr staging, showed there were significantly fewer MSA patients in the most severe stage (P = 0.024 and P = 0.048). Likewise, strata difference in scores for the SEADL indicated less dependency for MSA than for the PSP patients (P = 0.017). As expected, PSP patients scored worse on cognitive functions as assessed with the MMSE (P = 0.0001) (Table 2).

Table 5 Predictive validity of NNIPPS clinical diagnostic criteria in deceased patients with neuropathological diagnosis

<table>
<thead>
<tr>
<th></th>
<th>CGI&lt;4 n = 39</th>
<th>CGI ≥ 4 n = 73</th>
<th>ALL n = 112</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSP</td>
<td>MSA</td>
<td>PSP</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>0.95 (0.82–0.99)</td>
<td>0.88 (0.73–0.95)</td>
<td>0.95 (0.86–0.98)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>0.84 (0.7–0.89)</td>
<td>0.91 (0.79–0.96)</td>
<td>0.83 (0.74–0.87)</td>
</tr>
<tr>
<td>Overall fraction correct (95% CI)</td>
<td>0.90 (0.77–0.94)</td>
<td>0.90 (0.77–0.96)</td>
<td>0.89 (0.80–0.93)</td>
</tr>
<tr>
<td>Positive likelihood (95% CI)</td>
<td>6.01 (2.83–8.60)</td>
<td>9.71 (3.50–26.48)</td>
<td>5.68 (3.35–7.71)</td>
</tr>
</tbody>
</table>

Sensitivity, specificity, overall fraction correct and positive likelihood (95% CI) of NNIPPS diagnostic criteria in the overall population with neuropathology diagnosis, and broken down by disease severity as defined by the CGI. Early patients were defined as those below the median (CGI, range 1–3); late patients were those equal or over the median (CGI, range 4–6).
Demographics and clinical factors influencing survival were assessed with univariate and multivariate Cox model analysis (Supplementary material and Supplementary Table 7). In the univariate analyses, only variables scoring disease severity were predictive of survival. Accordingly, the stepwise multivariate analysis selected first the SEADL, followed by the CGI scales for disease severity and for dysautonomia, as strong predictors of survival.

However, following adjustment on these variables, disease duration was selected in the model (RR = 0.923, P = 0.007) indicating that at constant severity patients with a shorter disease duration at entry had a worse prognosis (‘fast progressors’). To visualise the discriminating accuracy of the combined variables, we constructed a prognostic score for each patient using a Cox coefficient weighted linear combination of the selected prognostic variables (Supplementary Fig. 1A). Following adjustment on the prognostic score as a variable, the strata factor became significant (RR = 0.657, P = 0.0002) indicating that PSP patients on the whole had a worse prognosis compared to MSA patients at constant disease severity and disease duration at entry (Supplementary Fig. 1B). Testing for a treatment effect on survival following adjustment on the prognostic variables did not change the results of the log-rank analysis nor was there any significant interaction between treatment and prognostic variables (Supplementary Table 3).

Discussion

Our results show that riluzole (up to 200 mg/day) is unlikely to be effective as a disease-modifying agent in PSP or MSA. Riluzole has a definite effect in ALS with an estimated 30–40% decrease in relative risk at 12–18 months (Bensimon et al., 1994; Lacomblez et al., 1996; Miller et al., 2007). NNIPPS was therefore designed to detect an effect of similar size. Power calculation assumptions for the placebo group were based largely on retrospective data (Litvan et al., 1996a, c, 2003; Testa et al., 1996, 2001; Ben-Shlomo et al., 1997), which underestimated the number of events observed in NNIPPS by 2–6%. Hence the NNIPPS trial achieved the power necessary to detect the hypothesized drug effect. The failure of NNIPPS to detect that effect could be due to heterogeneity in the trial population (loss of power), uneven distribution of demographic and clinical prognostic factors or comorbidity across treatment groups (randomization bias), poor compliance, or simply a smaller pharmacological effect in these conditions than the one used at the planning stage (effect size).

With regard to disease heterogeneity, a major difficulty in designing trials of disease-modifying therapy in neurodegenerative diseases is lack of diagnostic confidence, particularly at a stage of disease evolution when intervention is likely to be most effective. Indeed, there are no prospective studies of clinical diagnostic criteria assessed against the definitive criterion of histopathology in PSP or MSA (Litvan et al., 2003). Such studies are not easy to achieve, as long term follow-up and a high autopsy rate are required. Until validated biomarkers are available for use in trials, clinical criteria prospectively tested against pathology are essential to understand heterogeneity, which is a potential source of bias in estimating drug effects in Parkinson plus syndromes, as idiopathic Parkinson’s disease. Furthermore, criteria for large scale clinical trials in rare diseases such as PSP and MSA should be robust and simple, so that they can be applied in international multi-centre studies and not only in highly specialized centres. The NNIPPS diagnostic criteria represent a simplification of existing diagnostic criteria and we have shown them to have excellent convergent and predictive validities in this study population,
Despite variations in practice across three European countries and 44 centres. Although pathological examination was only possible in 112 cases (15% of recruited patients), the pathological sample is representative of those patients who died (45%) and, most likely, of the population as a whole. Further analysis of these diagnostic criteria in relation to MRI, DNA and neuropsychology is ongoing. At this stage it can be concluded that the estimated rate of misdiagnosis (about 6% overall, representing at most 46 patients in the ITT) is acceptably low and is unlikely to have biased the estimate of the drug effect. We could not detect imbalance in entry parameters with relevant prognostic significance for survival or disease progression which could account for the lack of treatment efficacy.

It is likely that a neuroprotective agent should be started in the earliest phase of the disease in order to demonstrate efficacy. Hence the impact of a neuroprotective agent is likely to be an inverse function of the stage of disease progression. In support of this assumption, riluzole did not show efficacy in ALS when tested in a trial including late stage disease patients (Bensimon et al., 2002). In the present trial, disease severity in the population included was greater than anticipated and the death rate was also higher than expected. Overall, about 50% of patients were classified as severely disabled or wheelchair bound (i.e. in the latest stages) by the Hoehn and Yahr staging and over 50% similarly classified as markedly, severely or extremely ill by the CGI scale for disease severity. However, adjustment for prognostic factors with the Cox model did not show any trend for a treatment effect, nor was there a significant treatment by prognostic factor interaction, meaning that the treatment effect is constant across the levels of the prognostic variables.

Even though treatment withdrawal rate was relatively high (25%), and might have had an impact on the size of a treatment effect, the overall mean dose was close to the maximum, and compliance rate was good (81.2%) considering the length of the trial, with a mean follow-up time close to the maximum planned (3 years). Finally, the PPPT sensitivity analysis was consistent with the ITT. Hence, the observed lack of treatment effect in this Parkinson plus population seems relatively robust. Our cohort clearly indicate that patients presenting with an akinetic-rigid syndrome and categorized as having PSP or MSA with rigid syndrome and categorized as having PSP or MSA with a primary endpoint in phase III studies and that our strategy of recruiting patients with PSP and MSA into a single stratified trial is methodologically valid where a generic neuroprotective effect is hypothesized. On the other hand, where the agent to be tested is thought to act on disease-specific pathways (e.g. processing of tau in PSP, or alpha-synuclein in MSA) a different strategy is likely to be more appropriate. The demographic features of the NNIPPS cohort clearly indicate that patients presenting with an akinetic-rigid syndrome and categorized as having PSP or MSA with reasonably high confidence have very poor prognosis, and no significant response to levodopa therapy. Unfortunately, there are no comparable prospective studies with which to compare our cohort. The European MSA Study Group (EMSA-SG) has reported preliminary data on a functional scale developed for MSA based on assessments of 50 patients diagnosed on the basis of the Gilman criteria (Gilman et al., 1998). A total of 412 patients with presumed MSA were included in a European registry (Geser et al., 2005, 2006), but no information on survival is yet available from the EMSA database. Nevertheless, demographic features of the 50 patients reported from the EMSA are broadly comparable to those in the NNIPPS MSA stratum (Geser et al., 2006). In a cohort study of 162 PSP patients the median survival since disease onset was 7.3 years, compared to 7.8 years in the NNIPPS cohort (Golbe and Ohman-Strickland, 2007). As with NNIPPS, functional scores predicted survival. However, multivariate analysis of prognostic factors revealed that age of onset and gender were predictors of survival. For the former, the apparent discrepancy is likely to be related to differences in the method of survival analysis since when survival is calculated from disease onset in our cohort, age of onset becomes a significant predictor of survival as well (data not shown). In relation to a gender effect, the sex ratio in most studies, as in the NNIPPS cohort, shows an equal proportion of men and women or a slight excess of men (Golbe et al., 1988; Santacruz et al., 1998; Schrag et al., 1999; Nath et al., 2001, 2003), whereas in the study of Golbe and Ohman-Strickland (2007) the sex ratio was reversed, indicating because they originate from selection of a biased sub-group. In early patients, when the likelihood of informative censoring is least, slope of change in functional measures might be appropriate as an end point in a definitive phase III trial where using survival as an endpoint would result in substantially increasing the number of patients and/or length of the trial. The parameters provided in this article allow power calculations for either context since linearity is a basic assumption for slope of change and should hold true early or late in the course of disease.

Third, survival data provide a unique insight into the natural history of these disorders and help to validate functional assessments in relation to prognosis. In this study, the median survival from onset of symptoms is in keeping with our prior assumptions based on retrospective (Litvan et al., 1996d; Testa et al., 1996, 2001; Vanacore et al., 2001a, b; Litvan et al., 2003) and prospective studies (Schrag et al., 2008). Following a step-wise Cox model adjustment on prognostic factors which selected disease severity scores (SEADL, CGI-ds, CGI-dysautonomia) and disease duration, there was a significant difference between strata consistent with the difference in the rate of disease progression between the two strata. Finally, our data show that it is feasible to use survival as a primary endpoint in phase III studies and that our strategy of recruiting patients with PSP and MSA into a single stratified trial is methodologically valid where a generic neuroprotective effect is hypothesized. On the other hand, where the agent to be tested is thought to act on disease-specific pathways (e.g. processing of tau in PSP, or alpha-synuclein in MSA) a different strategy is likely to be more appropriate. The demographic features of the NNIPPS cohort clearly indicate that patients presenting with an akinetic-rigid syndrome and categorized as having PSP or MSA with reasonably high confidence have very poor prognosis, and no significant response to levodopa therapy. Unfortunately, there are no comparable prospective studies with which to compare our cohort. The European MSA Study Group (EMSA-SG) has reported preliminary data on a functional scale developed for MSA based on assessments of 50 patients diagnosed on the basis of the Gilman criteria (Gilman et al., 1998). A total of 412 patients with presumed MSA were included in a European registry (Geser et al., 2005, 2006), but no information on survival is yet available from the EMSA database. Nevertheless, demographic features of the 50 patients reported from the EMSA are broadly comparable to those in the NNIPPS MSA stratum (Geser et al., 2006). In a cohort study of 162 PSP patients the median survival since disease onset was 7.3 years, compared to 7.8 years in the NNIPPS cohort (Golbe and Ohman-Strickland, 2007). As with NNIPPS, functional scores predicted survival. However, multivariate analysis of prognostic factors revealed that age of onset and gender were predictors of survival. For the former, the apparent discrepancy is likely to be related to differences in the method of survival analysis since when survival is calculated from disease onset in our cohort, age of onset becomes a significant predictor of survival as well (data not shown). In relation to a gender effect, the sex ratio in most studies, as in the NNIPPS cohort, shows an equal proportion of men and women or a slight excess of men (Golbe et al., 1988; Santacruz et al., 1998; Schrag et al., 1999; Nath et al., 2001, 2003), whereas in the study of Golbe and Ohman-Strickland (2007) the sex ratio was reversed, indicating
a potential sample bias. However, in MSA, demographic features such as age of onset and sex ratio are similar in the NNIPPS cohort to those recorded in previous natural history studies (Bower et al., 1997; Schrag et al., 1999; Vanacore et al., 2001a, b; Schrag et al., 2008).

In order to achieve a relatively homogeneous population of patients in the PSP strata, oculomotor abnormalities were mandatory at inclusion. This increased the accuracy of diagnosis (as confirmed pathologically) but did not take into account ‘PSP-Parkinsonism’ patients without oculomotor abnormalities (Williams et al., 2005, 2007) who may be confused with patients with PD or with MSA-P. Likewise, in order to recruit MSA patients with relatively early disease, we took a pragmatic approach so as to avoid, as far as possible, recruiting patients with disorders such as ozoopontocerebellar atrophy (Berciano et al., 2006) and idiopathic late-onset cerebellar ataxia, (Gilman and Quinn, 1996; Gilman et al., 2000). We expected this to reduce the proportion of patients with the cerebellar form of MSA (MSA-C) in our cohort, but cerebellar features were noted at presentation in 20% of our patients, similar to the findings of a prospective study (Schrag et al., 2008). Thus the NNIPPS population is reasonably representative of the whole (European and North American) MSA population. In this context, we have clearly shown that clinical features that have been regarded as characteristic of each disorder, and which are of key importance in the different consensus diagnostic criteria (Litvan et al., 1996a, b, d, 1997, 2003; Gilman et al., 1998) are common to both disorders. Thus at entry urinary symptoms were present in over 85% of patients in the MSA stratum but also in 50% in the PSP stratum (Fig. 5) and oculomotor signs, required for inclusion in the PSP stratum, were recorded in about 20% of patients in the MSA stratum. However, this included any evidence of supranuclear ophthalmoplegia, whereas allocation to the PSP stratum required significant supranuclear impairment of downward gaze. Likewise, behavioural and cognitive abnormalities were judged as being common in both PSP and MSA, in keeping with studies on cognitive function in these disorders (Robbins et al., 1992, 1994; Burk et al., 2006; Lyoo et al., 2008). Despite this overlap, we have shown that PSP and MSA can be differentiated by neurologists at a relatively early stage in the disease process using the NNIPPS diagnostic criteria.

A striking feature of the NNIPPS diagnostic criteria is the high sensitivity at inclusion, although a direct comparison with published diagnostic criteria is not possible as the timing of assessments was different (Osaki et al., 2002, 2004; Litvan et al., 2003). The NNIPPS diagnostic assessments were carried out on average 2 years before death in the pathologically confirmed cases whereas in the published criteria the assessments were made at the first (‘diagnostic’) visit when sensitivity is low, and at the end of the disease (i.e., on the last visit before death) when sensitivity is high. Studies of the sensitivity of the NNIPPS criteria at the earliest stages of disease process are now needed, though we provide evidence that the criteria still display acceptable quality in the sub-group below the mid-range of disease progression. Indeed, only because we studied both conditions in one cohort could we discern the close similarities and the diagnostically important differences between PSP and MSA.

We have shown that a large scale randomized trial with survival as the primary endpoint is feasible using the Parkinson plus concept to allow relatively early diagnosis and recruitment of a surprisingly homogeneous population, as shown by the performance of the NNIPPS diagnostic criteria informed by pathology. The rate of death was higher than anticipated, allowing us to shorten the trial slightly. We have learnt from clinical trials in ALS that survival does not necessarily equate with functional change (Lacomblez et al., 1996; Meininger, 2005). This has resulted in recommendations by the regulatory authorities (e.g. European Medicines Agency, Committee for Proprietary Medicinal Products, CPMP/ EWP/565/98) and by the World Federation of Neurology Committee on Research (Miller et al., 1999) that survival should be the primary endpoint for definitive demonstration of a neuroprotective drug effect. Future phase III studies on potential neuroprotective agents in these disorders should consider using survival as a primary endpoint, and combining patients with presumed PSP and MSA in the same trial in order to improve our understanding of both.

Supplementary material

Supplementary material is available at Brain online.

Funding

NNIPPS was an academic-led study with core funding from the European Union 5th Framework Programme (QLG1-CT-2000-01262); support from French Health Ministry, Programme Hospitalier de Recherche Clinique (AOM97073, AOM01125) and from Sanofi-Aventis affiliates in the UK, France and Germany providing an unconditional research grant and drug supply throughout the study. Three academic institutions (Institute of Psychiatry, King’s College London; Assistance Publique-Hôpitaux de Paris; and University of Ulm) were sponsors of the study in each country, and jointly own the data. The protocol and amendments were reviewed and approved by the Comité de Protection des Personnes of Pitié-Salpêtrière Hospital (France), the UK Multicentre Research Ethics Committee (MREC), (UK), Ethikkommission of the University of Ulm, (Germany), and by local Institutional Review Boards (Ethics Committees) where appropriate (UK, Germany).

Acknowledgements

We thank the patients and their families for their commitment and altruism, and The French and UK PSP Associations and the UK Parkinson’s Disease Research Group for their help and support. We are grateful to the many colleagues who were not formally part of the NNIPPS consortium but whose support contributed to the success of the study. The study protocol was filed in the open clinical trial registry (www.clinicaltrials.gov) with ID number NCT00211224.
References


Neuro-Imaging: Chair: M. Verin (Rennes, France), Members: N. Deasy, J. Jarosz (London, UK), T.K. Hauser (Tübingen, Germany), E. Kraft (Ulm, Germany), E. Broussolle (Lyon, France), D. Dormont, C. Marsault, A. Tourbah (Paris, France), L. Defevbre, L. Delmaire (Lille, France), Y. Roland (Rennes, France)

Neuro-Pathology: Chair: J.J. Hauw (Paris, France), Members: C. Duyckaerts, D. Seilhean (Paris, France), S. Al-Sarraj, T. Revesz (London, UK), B. Landwehrmeyer (Ulm, Germany), H. A. Kretzschmar (Munich, Germany)

Neuropsychology: Chair: R. Brown (London, UK), Members: T. Bak (Cambridge, UK), A. Danek (Munich, Germany), B. Dubois, L. Lacomblez (Paris, France), RM. Marié (Caen, France), I. Uttner (Ulm, Germany)

Genetics: Chair: A. Dür (Paris, France), Members: A. Al-Chalabi, N. Wood (London, UK), A. Bricé (Paris, France), W. Camu (Montpellier, France), K. Morrison (Birmingham, UK)

Logistics, Treatments, Monitoring, Data Management & Statistical analysis


Investigators within Countries

Principal Investigator France/Germany/UK

Centres (number of patients), Principal Investigators, Co-investigators (clinician/radiologist/psychologist)

France

Principal Investigator France: Y. Agid (Paris, France)

Aix en Provence (n = 20): F. Viallet, C. Couratier, S. Arguille (clinicians), H. Payan-Cassin, G.M. Vassault (radiologists), P. Henon, S. Gimeno (psychologists); Angers (5): F. Dubas, C. Fressinaud (Clinicians), J.Y. Tanguy (radiologist), D. Legall (psychologist); Besançon (n = 7): L. Rumbach, E. Vidry (clinicians), J. Kraehenbühl, J.F. Bonneville (radiologists), G. Chopard (psychologist); Caen (n = 13): F. LeDoze, G. Defer, F. Vialdec, R.M. Marié (clinicians), H. Huet (radiologist), F. Daniel, C. Lalevée (psychologists); Clermont-Ferrand (n = 21): F. Durif, B. Debilly, Ph. Derost, C. Tilignac (clinicians), J. Gabrillargues (radiologist), D. Lauvergne Crégu, C. Rosière (psychologists); Grenoble (n = 7): G. Besson, C. Mallaret (clinicians), S. Grand (radiologist), H. Klinger, A. Funkiewiez (psychologists); Lille (n = 8): A. Destée, L. Defevbre (clinicians), C. Delmaire (radiologist), K. Dujardin (psychologist); Limoges (n = 12): P. Courrèges (clinician), MP Bonconeur-Martel (radiologist), M. Chazot-Balcon (psychologists); Lyon (n = 11): E. Broussolle, H. Mollion (clinicians), M. Hermier (radiologist), M. Bouvard (psychologists); Marseille (n = 15): J.P. Azulay, T. Witjas (clinicians), (radiologist cf Aix en Provence), M. Delfini (psychologist); Montpellier (n = 16): W. Camu, F. Portet, J. Khoris, N. Pageot, G. Garrigues (clinicians), B. Viaud (radiologist), K. Martin, J. Bernard (psychologists); Nice (n = 8): M. Borg (clinician), S. Chanalet (radiologist), B. Ballet (psychologist); Paris patient clinical selection: M. Vidalhét, S. Sangla (Hôpital St Antoine), D. Ranoux (Hôpital St Antoine), J.P. Brandel (Hôpital Leopold Belland), T. De Broucker (Hôpital St Denis), Y. Agid, B. Dubois, Meininger, Verny (Hôpital Pitié-Salpêtrière), P. Cesaro (Hôpital Henri Mondor), G. Fenelon (Hôpital Tenon); Paris CIC Pitié-Salpêtrière-inclusion and follow-up (n = 111): Y. Agid, F. Bloch, A.M. Bonnet, L. Lacomblez, D. Maltière, A. Memin, Torni, M.L. Welter, J. Worbe (Clinicians), T. Lalain, A. Tourbah, C. Marsault, Pr. D. Dormont (Radiologists), B. Pillon, V. Czeremecki, A. Picard, B. Passaquet (psychologists); Pointe à Pitre (n = 5): D. Caparrós-Lefebvre, A. Lannuzel (clinicians), (no radiologist), F. Verlut (psychologist); Poitiers (n = 12): R. Gil, M. Baillé, S. Venisse, H. Moumy, V. Mesnage, J.L. Houeto, F. Petit (clinicians), P. Vandermarcq (radiologist), V. Bonnau, C. Ornon (psychologists); Rennes (n = 13): M. Verin (clinician), Y. Rolland (radiologist), P. Trébon, G. Salicé (psychologists); Strasbourg (n = 16): C. Tranchant, G. Steinmetz (clinicians), J.L. Dietemann (radiologist), Crémel (psychologist); Toulouse (n = 8): O. Rascol, M. Galitzky, C. Thalamas (clinicians), P. Manefle (psychologists), S. Lemcol, M.C. Deneuville, H. Delabaere (psychologists); Tours (n = 11): C. Prunier, A. Autret, P. Corsia (clinicians), P. Cottier, S. Gallas (radiologists), D. Beauchamp (psychologist).

Germany

Principal Investigator Germany: A. Ludolph (Ulm, Germany)


**United Kingdom**

**Principal Investigator UK:** PN Leigh (London, UK)

Belfast (n = 9): M. Gibson, R. Forbes (clinicians), C. Reynolds, C. S. McKinstry (radiologists), A. Dick (psychologist); Birmingham – City Hospital (n = 5): C. Clarke (clinician), S. Chavda (radiologist), S. Dhariwal, R. Hornabrook, S. Colhoun, V. Barnfield (psychologists); Birmingham – Queen Elizabeth Hospital (n = 9): H. Pall, D. Nicholls (clinicians), S. Chavda (radiologist), D. Nicholls (psychologist); Cambridge (n = 17): J. Hodges, T. Bak (clinicians), A. Carpenter (radiologist), T. Bak, V. Hearn, L. Donald (psychologists); Liverpool (n = 17): P. Moore (clinician), T. Dixon (radiologist), G. Baker, L. Owen, I. O’Brien (psychologists); London, King’s College London (n = 55): P.N. Leigh, K. R. Chaudhuri, D. Heaney, C. Blain, S. Azam, V. Williams, J. Iscacs, C. Smallman, B. Stanton (clinicians), J. Jarosz, N. Deasey (radiologists), R. Brown, A. Dittner, C. Donnellan, D. Secker, A. Langman, C. Lomax (psychologists); T. Simeon (research nurse) London NHNN & Queen Square Hospital (n = 29): A. Lees, N. Quinn, A. Evans, T. Scaravilli, N. Russo, E. Trikouli, D. Paviour, Luke Massey (clinicians), C. Andrews, J. Stevens (radiologists), M. Jahanshahi, D. Winterburn (psychologists); Middlesborough (n = 7): P. Newman, Bathgate (clinicians), N. Braden (radiologist), Z. Cowen (psychologist); Newcastle upon Tyne (n = 20): D. Burn, A. Zemansky, N. Warren (clinicians), P. English, A. Ghokhar (radiologists), J. Welch, P. Welch (psychologists); Stafford (n = 9): B. Summers (clinician), D. Steventon (radiologist), B. Summers, L. Silver (psychologists); Aberdeen (n = 4): C. Counsell (clinician), A. Murray (radiologist), J. Gordon, S. Harris, K. Perkins (psychologists); Guernsey (n = 9) S. Bhaumick, S. Evans, G. Turner (clinicians), (no radiologist), S. Adam (psychologist); Swansea (14): R. Weiser, C. Lawthom, A. Lowman (clinicians), (no radiologist), G. Forwood, M. Moran, L. Bastin (psychologists).