Movement Disorders
Vol. 22, No. 16, 2007, pp. 2371–2377
© 2007 Movement Disorder Society

Potential Outcome Measures and Trial Design Issues for Multiple System Atrophy

Susanne May, PhD,[1,2]* Sid Gilman, MD, FRCP,[3] B. Brooke Sowell, MS,[1] Ronald G. Thomas, PhD,[1,2] Matthew B. Stern, MD,[4] Amy Colcher, MD,[4] Caroline M. Tanner, MD, PhD,[5] Neng Huang, MD,[5] Peter Novak, MD, PhD,[6] Stephen G. Reich, MD,[7] Joseph Jankovic, MD,[8] William G. Ondo, MD,[8] Phillip A. Low, MD,[9] Paola Sandroni, MD,[9] Axel Lipp, MD,[9] Frederick J. Marshall, MD,[10] Frederick Wooten, MD,[11] Clifford W. Shults, MD,[2,12] and the North American Multiple System Atrophy Study Group

[1]*Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, California, USA*
[2]*Department of Neurosciences, University of California, San Diego, La Jolla, California, USA*
[3]*Department of Neurology, University of Michigan, Ann Arbor, Michigan, USA*
[4]*Parkinson's Disease and Movement Disorders Center, Pennsylvania Hospital, Philadelphia, Pennsylvania, USA*
[5]*Parkinson's Institute, Sunnyvale, California, USA*
[6]*Department of Neurology, Boston University, Boston, Massachusetts, USA*
[7]*Department of Neurology, University of Maryland, School of Medicine, Baltimore, Maryland, USA*
[8]*Department of Neurology, Baylor College of Medicine, Houston, Texas, USA*
[9]*Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA*
[10]*Department of Neurology, University of Rochester, Rochester, New York, USA*
[11]*Department of Neurology, University of Virginia Health System, Charlottesville, Virginia, USA*
[12]*Veterans Affairs San Diego Healthcare System, San Diego, California, USA*

Abstract: Multiple system atrophy (MSA) is a neurodegenerative disorder exhibiting a combination of parkinsonism, cerebellar ataxia, and autonomic failure. A disease-specific scale, the Unified Multiple System Atrophy Rating Scale (UMSARS), has been developed and validated to measure progression of MSA, but its use as an outcome measure for therapeutic trials has not been evaluated. On the basis of twelve months of follow-up from an observational study of 67 patients with probable MSA, we evaluated three disease-specific scores: Activities of Daily Living, Motor Examination, and a combined score from the UMSARS and two general health scores, the Physical Health and Mental Health scores of the SF-36 health survey, for their use as outcome measures in a therapeutic trial. We discuss related design issues and provide sample size estimates. Scores based on the disease-specific UMSARS seemed to be equal or superior to scores based on the SF-36 health survey. They appeared to capture disease progression, were well correlated and required the smallest sample size. The UMSARS Motor Examination score exhibited the most favorable characteristics as an outcome measure for a therapeutic trial in MSA with 1 year of follow-up. © 2007 Movement Disorder Society

Key words: Parkinsonism; power, study design; UMSARS; SF-36.

Multiple system atrophy (MSA) is a neurodegenerative disorder expressing a combination of autonomic failure, parkinsonism, and cerebellar ataxia. Disease progression is typically inexorable. The cause of MSA is unknown. Treatment options are only partially effective, for only a few symptoms.

The Unified Multiple System Atrophy Rating Scale (UMSARS) is a validated, disease-specific scale representing the diverse signs and symptoms in MSA.[1] It can assess rates of progression and is sensitive to change over time.[2] No instrument has been investigated for potential use in MSA clinical trials. Confidence in the clinical diagnosis of MSA is highest in those with "prob-

This manuscript is dedicated to the memory of Clifford W. Shults, MD.

Members of "The North American Multiple System Atrophy Study Group" are listed as an Appendix.

*Correspondence to: Susanne May, University of California San Diego, Division of Biostatistics and Bioinformatics, 9500 Gilman Dr.-M/C 0717, La Jolla, CA 92093-0717. E-mail: smay@ucsd.edu

Received 30 November 2006; Revised 5 August 2007; Accepted 16 August 2007

Published online 3 October 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/mds.21734

able MSA."[3] We investigated potential outcome measures for use in therapeutic trials of MSA, limiting investigation to those with "probable MSA". Three measures were derived from the disease-specific UMSARS: the Activities of Daily Living score (UMSARS-ADL, 12 questions), the Motor Examination score (UMSARS-ME, 14 questions) and a combined score (UMSARS-ADL + ME, 26 questions). Two measures from the SF36 health survey,[4] a standardized and validated instrument, were assessed: the Physical Health (SF36-PH, 21 questions) and Mental Health (SF36-MH, 14 questions) scores. Poorer health is signified by lower scores on the SF-36 and by higher scores on the UMSARS scales. The five outcome measures were compared to determine (a) whether they capture the decline in health, (b) whether they exhibit high correlation over time and (c) which measure requires the smallest sample sizes (smallest signal-to-noise ratio).

METHODS

Subjects and Evaluation

We studied participants currently enrolled at nine US movement disorder centers in an observational and risk factor study of MSA.[5] Patients were followed biannually. All centers obtained Institutional Review Board approval. All subjects provided informed consent. All subjects met Consensus Criteria for probable MSA.[3] Each investigator reviewed an UMSARS training video prior to enrolling patients. Sixty-seven patients completed an initial and 12 month evaluation and form the basis of this report.

Patients were classified by MSA subtype using the Consensus Criteria within the limits of available information.[3] As patients' initial study evaluation was on average 4.6 years after diagnosis, it was not possible to determine whether parkinsonian or cerebellar symptoms predominated at the onset of the disease for some patients. Rather than force an arbitrary classification, we have chosen to consider these subjects with both parkinsonian and cerebellar criterion at enrollment but for whom predominant criterion at onset was unclear, as "mixed". Health decline was estimated separately for these "mixed" patients to allow detection of differences between this group of patients and patients who could unambiguously be designated as MSA-P or MSA-C. As a result, the initial study examinations, medical records and, as needed, information from the treating physician was used to determine MSA subtype in the following way:

i. MSA-Parkinsonism (MSA-P): Patients meeting parkinsonian but not cerebellar criterion and patients for whom parkinsonism preceded cerebellar signs by at least 1 year.
ii. MSA-Cerebellar (MSA-C): Patients meeting cerebellar but not parkinsonian criterion and patients for whom cerebellar signs preceded parkinsonism by at least 1 year.
iii. MSA-Mixed (MSA-M): Patients with parkinsonism and cerebellar criterion at enrollment for whom the presenting signs could not be determined.

Because of disease progression, 10 of the 67 patients (15%) could not attend the 12 month visit. Study neurologists obtained UMSARS-ADL scores by telephone. The other measures could not be assessed.

Statistical Analysis

Summary statistics were presented as means and standard deviations (SD). Comparisons of initial evaluation and month 12 measures were based on Wilcoxon signed rank tests. Spearman and Pearson estimates of correlation were determined. Wilcoxon rank sum tests and Kruskal–Wallis tests were used to assess whether change scores depended on patient characteristics. Continuous factors (like age) were dichotomized at the median for these analyses. Normality of change scores were assessed using Shapiro–Wilk tests. Observed change scores, correlations over time and the larger of the standard deviations at the two time points provided the basis for sample size calculations. ANCOVA models adjusting for baseline scores were used to estimate sample sizes[6] using two-sided tests and alpha levels of 0.05. The formula for the sample size is: $n = 2\sigma^2/\delta^2[1 - \rho^2] \times (z(\alpha/2) + z(\beta))^2$, where σ is the standard deviation, δ is the absolute difference observed at the pre- and post-visit, ρ is the pre/post correlation, and $z(\alpha/2)$ and $z(\beta)$ are the z-scores of the standard normal distribution relating to the size and power of the test, respectively. Of note, this formula is simplified compared to the Frison and Pocock presentation, as the number of pre- and post-visits (r and p in Frison and Pocock's notation) both have a value of one. Data analyses were performed using the statistical software R and STATA.[7,8] P-values were not adjusted for multiple comparisons.

RESULTS

The majority of patients were non-Hispanic, Caucasian men. Education beyond high school was common (Table 1). Mean age at MSA onset and enrollment was 60.5 and 65.0 years, respectively. Most patients had MSA-P (60%); 27% MSA-M and 13% MSA-C.

Forty-seven of the 67 patients (70%) were taking a dopaminergic medication at their initial evaluation.

TABLE 1. *Patient characteristics at initial evaluation (N = 67)*

	N/% or mean (SD)
General	
Gender male (N/%)	40/59.7
Age [mean(SD)], yr	65.0 (9.2)
Marital status (N/%)	
Married	59/88.1
Widowed	2/3.0
Divorced	4/6.0
Never married	2/3.0
Race (N/%)	
Asian	4/6.0
Caucasian	62/92.5
Other or unknown	1/1.5
Ethnicity (N/%)	
Hispanic or Latino/a	2/3.0
Not Hispanic or Latino/a	65/97.9
Education [yr, Mean(SD)]	15.8 (2.8)
Disease-specific	
MSA type (N/%)	
MSA-parkinsonism	40/59.7
MSA-cerebellar	9/13.4
MSA-mixed	18/26.9
Age at onset of MSA [Mean(SD)]	60.5 (9.9)
Disease duration [yr, Mean(SD)]	4.6 (3.3)
Beneficial levodopa response (N/%)	
No	20/29.9
Yes	22/32.8
Not received	21/31.3
Initial but not continuing response	4/6.0
Disability scale* (N/%)	
Independent	3/4.5
Less independent	22/32.8
More dependent	15/22.4
Very dependent	23/34.3
Totally dependent	4/6.0

*Independent: Completely independent, able to do all chores with minimal difficulty or impairment, essentially normal, unaware of any difficulty. Less independent: Not completely independent, needs help with some chores. More dependent: Help with half of chores, spends a large part of the day with chores. Very dependent: Now and then does a few chores alone or begins alone, much help needed. Totally dependent: Totally dependent and helpless, bedridden.

Among these, 3 had discontinued any dopaminergic medication by their month 12 visit, but another 7 patients had started taking dopaminergic medication between their initial evaluation and their month 12 visit. Also, 32 patients (48%) were taking medication for postural hypotension at their initial evaluation, with 2 patients discontinuing and 5 patients starting medication for postural hypotension between their initial evaluation and month 12 visit. About one-third reported a sustained benefit from Levodopa at enrollment, but nearly equal numbers either had no Levodopa benefit (31%) or never received Levodopa (31%). For the 26 patients who had benefit from Levodopa, the average number of years of benefit was 3.3 (SD = 2.1). The average maximum Levodopa dose reported at the initial evaluation was 664.9 mg/day (SD = 453.5).

At enrollment, the majority of patients (63%) reported at least some dependency on others for their daily chores, and 6% were totally dependent. After 12 months, 78% reported at least some dependency for their daily chores [more dependent: n = 18 (27%); very dependent: n = 27 (40%); totally dependent: n = 7 (10%)].

Potential Outcome Measures

All of the potential outcome measures except for the SF36-MH score ($P = 0.09$) showed a significant decline (Wilcoxon signed rank tests all P-values < 0.001) in health (Table 2). For all scales, correlations between initial evaluation and 12 month follow-up were high and statistically significant (Spearman rank tests all P-values < 0.001). The percent of patients for whom the scores indicated a decline in health over time varied from 55% (SF36-MH) to 81% (UMSARS-ADL + ME). Of note, average UMSARS-ADL and UMSARS-ME change scores did not add up to average UMSARS-ADL + ME change scores, because UMSARS-ADL scores were available for more patients than UMSARS-ME scores.

A number of factors were evaluated regarding their influence on disease progression, including gender, age at enrollment, age at disease onset, education, MSA type, disease duration, disability status (UMSARS) and baseline scores. Decreases in SF36-MH scores were significantly larger for older patients ($P = 0.005$, age ≥ 66 versus <66) and patients who were diagnosed later in life ($P = 0.04$, age at onset ≥ 60 versus <60). Patients with higher baseline SF36-MH scores (indicating better health) experienced a significantly larger decline in SF36-MH scores ($P = 0.03$, baseline SF36-MH score $\geq$ 53.6 versus $<$53.6). Of note, similar relationships were borderline significant for the ME and SF36-PH Score. None of the other factors (including MSA type) showed a significant association with disease progression. Normality assumptions of the change scores were not significantly violated except for the SF36-PH scale ($P < 0.01$, exclusion of one outlier resulted in $P > 0.50$).

Required sample sizes for a potential therapeutic trial with 1 year follow-up were estimated for the UMSARS-ADL, UMSARS-ME, UMSARS-ADL + ME and SF36-PH scores as these showed a statistically significant decline in health (Table 3). Sample size estimates were determined for a 25, 33, and 50% reduction in decline. The required sample sizes per group ranged from 98 to 1,024 to achieve 90% power and from 73 to 765 to achieve 80% power. Specifically, 98 patients per group were required to detect with 90% power a difference of 2.25 absolute points (Table 3, row 6) for the UMSARS-ME score assuming that the UMSARS-ME

TABLE 2. *Characteristics of five scores measuring changes in health over time [except for ρ, results are presented as mean (standard deviation)]*

	N		Initial evaluation	Month 12	ρ*	Absolute change**	Percent change	N (%)*** indicating health decline
UMSARS-ADL	67	Mean (SD)	23.9 (7.2)	27.0 (8.1)	0.79	3.1 (5.0)	15.1 (24.6)	46 (68.7)
		Median (IQR)	24.0 (9.0)	28.0 (11.0)	0.75[a]	3.0 (6.0)[a]	13.3 (26.1)	
UMSARS-ME	57	Mean (SD)	24.0 (7.0)	28.5 (7.2)	0.74	4.5 (5.2)	23.8 (34.9)	45 (78.9)
		Median (IQR)	24.0 (10.0)	29.0 (8.0)	0.72[a]	4.0 (6.0)[a]	16.0 (21.0)	
UMSARS-ADL+ME	57	Mean (SD)	48.1 (12.4)	55.3 (13.2)	0.78	7.1 (8.5)	17.2 (23.2)	46 (80.7)
		Median (IQR)	48.0 (14.0)	55.0 (18.0)	0.78[a]	6.0 (10.0)[a]	13.0 (18.8)	
SF36-PH	56	Mean (SD)	36.5 (16.3)	29.8 (16.6)	0.71	−6.8 (12.6)	−15.9 (31.4)	37 (66.1)
		Median (IQR)	36.6 (22.6)	27.3 (17.6)	0.73[a]	−5.6 (16.9)[a]	−17.9 (49.5)	
SF36-MH	56	Mean (SD)	53.7 (15.8)	50.7 (15.8)	0.69	−3.0 (12.5)	−2.0 (29.8)	31 (55.4)
		Median (IQR)	53.6 (21.4)	50.0 (21.4)	0.66[a]	−1.8 (19.6)	−4.1 (35.7)	

UMSARS-ADL, Activities of daily living; UMSARS-ME, motor examination score; UMSARS-ADL+ME, sum of UMSARS-ADL and UMSARS-ME score; SF36-PH, SF-36 overall physical health score; SF36-MH, SF-36 overall mental health score.

*ρ, Pearson correlation of initial evaluation and 12 month scores.

**Determined as the average of the difference between the outcome measure value at month 12 and the outcome measure value at the initial evaluation.

***Determined as the number of patients with a score at month 12 that indicates worse health compared to the score at the initial evaluation.

[a]Spearman rank test *P*-value < 0.001.

score changes by 4.5 absolute points in the control group (from 24.0 at the initial evaluation to 28.5 at 12 months). In general, required sample sizes were above 150 per group for the SF36-PH and the UMSARS-ADL scale and were less than or close to 100 per group only for the UMSARS-ME and the combined UMSARS-ADL + ME scales if the observed effect represented a 50% reduction in decline over 12 months. As the UMSARS-ME score resulted in the smallest sample size per group, more detailed estimates for various percent reductions in decline are presented in Figure 1. For example, 203 patients per group (total sample size of 406) were required to achieve 80% power to detect a 30% reduction in UMSARS-ME score over 12 months.

As an example calculation estimating the sample sizes, consider the estimated sample size of $n = 73$ for the UMSARS motor examination for an effect size of 50% and 80% power. For this setting, $\delta = 2.25$, $\rho = 0.74$, $z(\alpha/2) = z(0.975) = 1.960$, $z(\beta) = z(0.80) = 0.842$, and $\sigma = 7.2$. As a result, $n = 2 \times 7.2^2/2.25^2\ [1 - 0.74^2]\ (1.960 + 0.842)^2 = 72.7$.

DISCUSSION

This study demonstrated that all scores, except for the SF36-MH score, reflected significant health decline in MSA patients. As all measures appeared well correlated over time, they were good candidates as outcome measures. A factor that strongly influences the required sam-

TABLE 3. *Sample size estimates (per group) for ANCOVA analyses adjusting for baseline levels*

	Month 12 control	Month 12 treatment	Diff* At 1 yr	Percent**	ρ***	Power	
						90%	80%
UMSARS-ADL		26.23	0.77	25		863	645
	27.0	25.98	1.02	33	0.79	496	370
		25.45	1.55	50		216	162
UMSARS-ME		27.38	1.13	25		390	291
	28.5	27.02	1.49	33	0.74	224	167
		26.25	2.25	50		98	73
UMSARS-ADL+ME		53.50	1.80	25		433	331
	55.3	52.92	2.38	33	0.78	254	190
		51.70	3.60	50		111	83
SF36-PH		31.48	−1.68	25		1024	765
	29.8	32.01	−2.21	33	0.71	588	439
		33.15	−3.35	50		256	192

*Defined as the difference after one year between the average value in the outcome measure in the treatment group and the average value in the outcome measure in the control group.

**Change in treatment group relative to change in control group.

***ρ, assumed correlation between initial evaluation and 12 month scores.

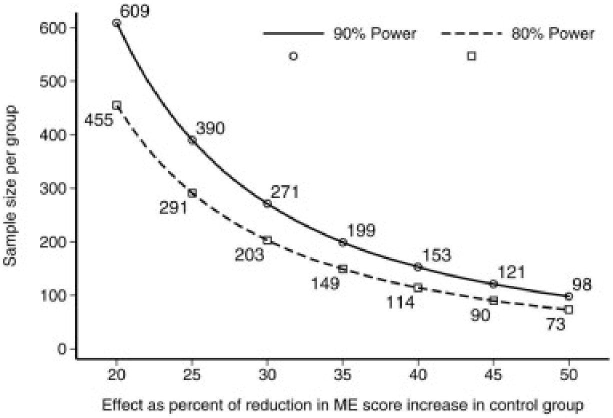

FIG. 1. Required sample size (per group) for various effect sizes for the Motor Examination score.

ple sizes is the amount of variability in a measure relative to its absolute values (signal-to-noise ratio). In comparisons of the five scales with respect to their ability to measure the decline in health over 12 months and in comparisons of estimated required sample sizes, four of the measures (UMSARS-ADL, UMSARS-ME, UMSARS-ADL + ME, and SF36-PH) were judged favorably by the first criterion and among these four the UMSARS-ME was most efficient by the second criterion, but closely followed by the combined UMSARS-ADL + ME score.

This is the largest prospective study thus far to examine outcome measures in MSA patients. The most relevant existing information is based on a mixed study population of possible and probable cases[2] or focuses on patients with MSA-parkinsonism.[9]

One strength of this study is that the study population consists entirely of patients with a diagnosis of probable MSA. In fact, this is the first study to contrast disease-specific and general scales with respect to their utility for a therapeutic trial that relied on these more stringent diagnostic criteria.

One of the limitations of this research is that a minimal clinically important change[10] has not been established for MSA patients and thus was unavailable as a reference point. The lack of significant differences in health decline for different MSA types and age, for example, could indicate either that there are no true significant differences or that the sample size was not large enough to rule out the possibility that chance accounted for the observed differences. In addition, differences in health decline for the classic MSA subtypes, MSA-P and MSA-C, might not have been observed because of progression of both, parkinsonian and cerebellar symptoms at the time of enrollment in some patients, designated here as MSA-M, and could reflect comparable health decline for later stages of the disease. Of note, if patients who were reported as mixed MSA type were randomly assigned to MSA-P and MSA-C in the same proportions as the designated patients (80% MSA-P and 20% MSA-C) and hypotheses tests involving MSA type were repeated in sensitivity analysis (data not shown), results and conclusions did not change. Also, improved estimates of disease progression could be achieved by a larger study population and longer follow-up. A longer follow-up time, however, has the potential to bias the results due to drop-out of patients who enter more advanced stages of the disease. Among the 67 patients, 6 have died and the diagnosis of MSA was confirmed by autopsy. For patients without this neuropathological confirmation there is a potential for misdiagnosis. Nevertheless, as only patients with probable MSA are included in this study, we expect the potential for misdiagnosis to be less than for a population of possible MSA patients. Another limitation of this research is it is not a population based study in that patients are recruited from the patient population of the study neurologists. As such, results might not be generalizable to all US based MSA patients.

Previous research has shown a faster disease progression for patients who had been diagnosed more recently.[2] This result was not confirmed here, but might be explained by the large percent of possible MSA patients in the previous study.

Considering only patients with a diagnosis of probable MSA for a potential therapeutic trial has disadvantages as well. Patients who are at an earlier stage of the disease may benefit most from a therapeutic agent.[2,11] This in turn may also reduce the number of patients that would need to be included in a trial. On the other hand, misdiagnosis of MSA would be more likely within this patient group, and this might diminish the benefits of a reduced sample size. To avoid this, the use of surrogate markers has the potential to assist such a study. These might include CSF markers to make the earliest possible diagnosis[12-15]; autonomic function tests such as the arginine growth hormone stimulation test[16]; and neuroimaging methods.[17-20] All of these approaches have the potential to improve the necessary sample size for a therapeutic trial in MSA. Accordingly, our estimates of a required sample size represent the most conservative scenario.

The choice of outcome measure for a therapeutic trial in MSA patients should consider other factors as well. For example, the severity of the disease might not allow patients to return to the clinic for a follow-up exam, but

UMSARS-ME scores require that a clinician examines the patient in person. Thus, for a therapeutic trial the possibility of home visits might need to be considered. Alternatively, the design might need to take into account differential drop-out if UMSARS-ME scores are used as the primary outcome measure and home visits are infeasible.

Randomization is standard for therapeutic trials and can be expected to balance groups with respect to patient characteristics when many trials are considered. For any individual trial, balance across groups cannot be guaranteed if simple or block randomization is used. Factors that show a potential for confounding treatment effects might be candidates for stratified or adaptive randomization. In this study, gender appeared to have a significant and borderline significant effect if disease progression was measured by the UMSARS-ADL and UMSARS-ADL + ME scores, respectively. If either of these is used as outcome measure, stratified or adaptive randomization should be considered with respect to gender. In addition, some of the measures showed significant or borderline significant differences in decline depending on baseline scores; thus, in addition to controlling for baseline scores analytically, stratified or adaptive randomization might also be considered for these.

The scores based on the disease-specific instrument (UMSARS) were equal or superior to the scores based on the SF-36 health survey for all three criteria that were used to evaluate the scales as potential outcome measures. This is not surprising as the UMSARS was designed to capture disease progression for a rare and complex disease, and confirms the utility of this relatively new instrument not only for observational studies, but also for therapeutic trials. However, considering the number of MSA patients that can realistically be recruited into a US based therapeutic trial, this study shows that potential interventions need to reduce the decline in health by at least 50%.

APPENDIX

The North American Multiple System Atrophy Study Group. Steering Committee—University of California at San Diego: Clifford W. Shults, MD, Eliezer Masliah, MD, Ron Thomas, PhD, Susanne May, PhD; University of Michigan: Sid Gilman, MD; Parkinson's Institute: Caroline Tanner, MD, PhD; University of Washington: Walter Kukull, PhD; University of Pennsylvania: Virginia Lee, PhD, John Trojanowski, MD, PhD; Mayo Clinic, Rochester: Phillip Low, MD; University of Rochester: Ira Shoulson, MD; Albert Einstein College of Medicine: Laurie Ozelius, PhD; Indiana University: Tatiana Foroud, PhD.

Study Investigators (not listed as coauthors): Thomas Chelimsky, MD, University Hospital of Cleveland.

Study Coordinators: Debra Berry, University of Rochester; Marsha Burks, University of Michigan; Nancy Zappala, University of Maryland; Toni Gehrking, Mayo Clinic; Melissa Diggin, Boston University; Ernesto Jimenez, Baylor College of Medicine; Kathleen Comyns, Parkinson's Institute; Mary Lloyd, Pennsylvania Hospital; Deborah Fontaine, University of California at San Diego.

Acknowledgments: The work was supported by NIH grant P01 NS044233 and the William M. Spencer, Jr. Research Fund. We thank the editor and referees for very helpful comments, which markedly improved the manuscript.

REFERENCES

1. Wenning GK, Tison F, Seppi K, et al. Development and validation of the unified multiple system atrophy rating scale (UMSARS). Mov Disord 2004;19:1391-1402.
2. Geser F, Wenning GK, Seppi K, et al. Progression of multiple system atrophy (MSA): a prospective natural history study by the European MSA study group (EMSA SG). Mov Disord 2006;21: 179-186.
3. Gilman S, Low PA, Quinn N, et al. Consensus statement on the diagnosis of multiple system atrophy. J Neurol Sci 1999;163:94-98.
4. Ware J, Kosinski M, Keller S. SF-36 physical and mental health summary scales: a user's manual, 4th ed. Boston, MA: The Health Institute, New England Medical Center; 1994.
5. Gilman S, May S, Shults CW, et al. The North American Multiple System Atrophy Study Group. J Neural Transm 2005;112:1687-1694.
6. Frison L, Pocock S. Repeated measurements in clinical trials: analysis using mean summary statistics and its implications for design. Stat Med 1992;11:1685-1704.
7. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2005. URL http://www.R-project.org.
8. StataCorp. Stata statistical software: release 9. College Station, TX: StataCorp LP; 2005.
9. Seppi K, Yekhlef F, Diem A, et al. Progression of Parkinsonism in multiple system atrophy. J Neurol 2005;252:91-96.
10. Schrag A, Sampaio C, Counsell N, Poewe W. Minimal clinically important change on the unified Parkinson's disease rating scale. Mov Disord 2006;21:1200-1207.
11. Watanabe H, Saito Y, Terao S, et al. Progression and prognosis in multiple system atrophy: an analysis of 230 Japanese patients. Brain 2002;125:1070-1083.
12. Brettschneider J, Petzold A, Sussmuth SD, et al. Neurofilament heavy-chain NfH(SMI35) in cerebrospinal fluid supports the differential diagnosis of Parkinsonian syndromes. Mov Disord 2006; 21:2224-2227.
13. Abdo WF, DeJong D, Hendriks JC, et al. Cerebrospinal fluid analysis differentiates multiple system atrophy from Parkinson's disease. Mov Disord 2004;19:238-240.

14. Abdo WF, van de Warrenburg BP, Munneke M, et al. CSF analysis differentiates multiple-system atrophy from idiopathic late-onset cerebellar ataxia. Neurology 2006;8:474-479.
15. Holmberg B, Johnels B, Blennow K, Rosengren L. Cerebrospinal fluid Aβ42 is reduced in multiple system atrophy but normal in Parkinson's disease and progressive supranuclear palsy. Mov Disord 2003;18:186-190.
16. Pellecchia MT, Longo K, Pivonello R, et al. Multiple system atrophy is distinguished from idiopathic Parkinson's disease by the arginine growth hormone stimulation test. Ann Neurol 2006;60: 611-615.
17. Nicoletti G, Lodi R, Condino F, et al. Apparent diffusion coefficient measurements of the middle cerebellar peduncle differentiate the Parkinson variant of MSA from Parkinson's disease and progressive supranuclear palsy. Brain 2006;1299: 2679-2687.
18. Seppi K, Scherfler C, Donnemiller E, et al. Topography of dopamine transporter availability in progressive supranuclear palsy: a voxelwise [123I]β-CIT SPECT analysis. Arch Neurol 2006;63: 1154-1160.
19. Seppi K, Schocke MFH, Prennschuetz-Schuetzenau K, et al. Topography of putaminal degeneration in multiple system atrophy: a diffusion magnetic resonance study. Mov Disord 2006;21:847-852.
20. Eckert T, Barnes A, Dhawan V, et al. PET in the differential diagnosis of parkinsonian disorders. Neuroimage 2005;26:912-921.