Original Article

Title: Comparison of different symptom assessment scales for multiple system atrophy

Authors: *Masaaki Matsushima, M.D.[1], Ichiro Yabe, M.D., Ph.D.[1], Koji Oba, M.D., Ph.D.[2, 3], Ken Sakushima, M.D., Ph.D.[1], Yasunori Mito, M.D.[4], Asako Takei, M.D., Ph.D.[5], Hideki Houzen, M.D., Ph.D.[6], Kazufumi Tsuzaka, M.D., Ph.D.[7], Kazuto Yoshida, M.D., Ph.D.[8], Yasunori Maruo, M.D., Ph.D.[9], Hidenao Sasaki, M.D., Ph.D.[1]

*Corresponding author

Affiliations: 1. Department of Neurology, Hokkaido University Graduate School of Medicine; 2. Research and Clinical Trial Center, Hokkaido University Hospital; 3. Department of Biostatistics, School of Public Health, Graduate School of Medicine, and Interfaculty Initiative in Information Studies, The University of Tokyo; 4. Sapporo City General Hospital; 5. Hokuyukai Neurological Hospital; 6. Obihiro Kosei Hospital; 7. Kushiro Rosai Hospital; 8. Japanese Red Cross Asahikawa Hospital; 9. Hakodate Municipal Hospital

Addresses: 1. Kita-15, Nishi-7, Kita-ku, Sapporo 060-8638, Japan; 2. 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; 3. Kita-14, Nishi-5, Kita-ku, Sapporo 060-8648, Japan; 4. 1-1 Kita-11, Nishi-13, Chuo-ku, Sapporo 060-8604, Japan; 5. Niju-Yon-Ken 2-2-4-30, Nishi-ku, Sapporo 063-0802, Japan; 6. Nishi-6, Minami-8, Obihiro 080-0016, Japan; 7. 13-23 Nakazono-cho, Kushiro 085-8533, Japan; 8. 1-1

Akebono, Asahikawa 070-8530, Japan; 9. 1-10-1 Minato-cho, Hakodate 041-8680, Japan

Corresponding author: Masaaki Matsushima

Tel: +81-11-706-6028

Fax: +81-11-700-5356

E-mail: mmasaaki@huhp.hokudai.ac.jp

Key words: multiple system atrophy; standardized response mean; symptom assessment scale; short-term change; sample size

Financial Disclosure: Hidenao Sasaki received a Grant-in-Aid (No 27280401) from Japan Agency for Medical Research and Development, a Grant-in-Aid from Takeda Science Foundation and a Grant-in-Aid for the Research Committee for Ataxic Diseases of the Research on Measures for Intractable Diseases from the Ministry of Health, Labour and Welfare, Japan.

The other authors have no financial disclosures to report.

The authors disclose no conflicts of interest.

None of the authors received funding for this study.

Abstract

To identify the most sensitive scale for use in clinical trials on multiple system atrophy (MSA), a short and sensitive scale is needed for MSA clinical trials. Potential candidates are the Unified MSA Rating Scale (UMSARS), Scale for the Assessment and Rating of Ataxia (SARA), Berg Balance Scale (BBS), MSA Health-Related Quality of Life scale (MSA-QoL), and Scales for Outcomes in Parkinson's Disease–Autonomic questionnaire (SCOPA-AUT). We enrolled patients with MSA from eight hospitals in Hokkaido, Japan. Board-certified neurologists assessed each patient at 6-month intervals and scored them on the UMSARS, SARA, BBS, MSA-QoL, and SCOPA-AUT. Score changes were evaluated using the standardized response mean (SRM). The correlation between disease duration and each score was examined. The first evaluation was conducted on 85 patients (60 patients with MSA cerebellar ataxia dominant subtype [MSA-C] and 25 patients with MSA Parkinsonism-dominant subtype [MSA-P]). Sixty-nine patients were examined after 6 months and 63 patients after 12 months. The UMSARS Part 4 had the largest SRM after 6 months and the SARA after 12 months. SRMs for MSA-P, the shorter duration group, and the early-onset group were larger than were those for MSA-C, the longer duration group, and the late-onset group. SRMs for items regarding skilled hand activities, walking, and standing were relatively large. Our study indicates that the UMSARS (parts 2 and 4), SARA, and BBS are sensitive scales for evaluating MSA progression over 12 months. Items with large SRMs effectively evaluated short-term changes.

Introduction

Multiple system atrophy (MSA) is an adult-onset, rapidly progressing neurodegenerative disease characterized by autonomic dysfunction, Parkinsonism, and ataxia [1, 2]. While the pathogenesis of MSA, such as α-synuclein aggregation [3], tubulin polymerization-promoting protein impairment [4], inflammatory mechanisms [5], mitochondrial dysfunction, and COQ2 mutation [6] has been revealed, and animal models [7] have been used in recent years, an effective treatment for MSA has not been realized. Thus, it is hoped that emerging treatments, such as selective serotonin reuptake inhibitors [8], intravenous immunoglobulin [9], mesenchymal stem cell therapy [10], and induced pluripotent stem cell therapy [11], will prove useful.

However, specific MSA biomarkers that can be measured through serum or cerebrospinal fluid analysis or imaging have not been yet identified. To facilitate the development of effective treatments, a short, sensitive symptom assessment scale is needed for use in large multicenter clinical trials on MSA, but such a scale has not been established.

The Unified MSA Rating Scale (UMSARS) [12] is a comprehensive scale for MSA assessment based on scales such as the Unified Parkinson's Disease Rating Scale (UPDRS) [13], International Cooperative Ataxia Rating Scale (ICARS) [14], and Composite Autonomic Symptom Scale (COMPASS) [15], which were developed for use in the European MSA (EMSA) study [16]. The UMSARS has been used in many clinical studies, such as the EMSA [17] and the North American MSA (NAMSA) study [18]. The

UMSARS measures various symptoms of MSA and is very useful for total symptom assessments of MSA. However, it typically takes about 20 min to complete all items, meaning that it is not a “brief” scale and is difficult to conduct quickly on MSA outpatients. Further, as several items of the UMSARS Part 1 (medical interviews) resemble those of the UMSARS Part 2 (motor examinations), certain items may be redundant. Therefore, the UMSARS may be superior to other scales for use in MSA clinical trials, but a shorter assessment may be better. The development of such a scale was the motivation for the present study.

In addition to the UMSARS, the Scale for the Assessment and Rating of Ataxia (SARA) [19] is a brief scale for evaluating ataxia. The SARA assesses symptoms specific to ataxia, and it has been shown to be equivalent to the ICARS [20]. The required time for the examination of SARA was approximately 4 min [20]. However, it is not sufficient for the evaluation of parkinsonism.

The Berg Balance Scale (BBS), developed in 1989, is a scale that measures balance disturbances [21]. The BBS has been used to assess balance disturbances associated with various conditions, such as stroke and orthopedic diseases. Several reports have mentioned the relationship between BBS and activities of daily living (ADL) in Parkinson’s disease [22]. Meanwhile, the BBS needs certain examination tools (e.g., ruler and stool).

SARA and the BBS were not validated in patients with MSA, but several studies have used SARA for evaluating symptoms of MSA [23]. BBS has a high reliability among neurodegenerative diseases such as

Parkinson's disease [24]. BBS is suitable for the evaluation of MSA because it can assess balance disturbances derived from not only ataxia but also Parkinsonism. Inter-rater reliability of the UMSARS, SARA, and BBS has been studied previously and found sufficiently reliable [12, 20, 24].

The MSA Health-Related Quality of Life scale (MSA-QoL) is a self-report assessment especially designed for MSA [25]. The motor subscale of the MSA-QoL has been demonstrated to be equivalent to the UMSARS in usefulness [26]. However, the MSA-QoL consists of more than 40 items (plus a visual analogue scale), which is somewhat inconvenient to patients and their caregivers.

The Scales for Outcomes in Parkinson's Disease–Autonomic questionnaire (SCOPA-AUT) is a self-assessment scale for autonomic dysfunction [27]. While one report used this scale for MSA evaluation [28], it was originally developed for Parkinson's disease. However, typically, the SCOPA-AUT does not adequately assess motor function. The COMPASS also evaluates autonomic dysfunction, but the UMSARS includes a part of the COMPASS, so here we used the SCOPA-AUT.

In a clinical trial of riluzole for Parkinson plus syndrome (progressive supranuclear palsy and MSA parkinsonism dominant subtype [MSA-P]), the NNIPPS scale, an original comprehensive scale using standardized response means (SRM), was employed [29]. While this scale might have been useful, it requires 30 to 40 min to complete the items. Thus, we did not consider the NNIPPS scale in this study. However, SRMs, which are an effect size estimating responsiveness, are thought to be suitable for comparing various different indices. A previous study used SRMs to compare the influences of five

different instruments [30], and with reference to such a precedent, we decided to use SRMs in this study.

To date, there has been no study that directly compares the sensitivity among different MSA scales. We considered the UMSARS, SARA, BBS, MSA-QoL, and SCOPA-AUT as potential candidates for the best scale for use in MSA clinical trials, while also looking at the role SRMs play, if any, with these scales.

Subjects and methods

Patients with MSA were recruited from eight different hospitals (Hokkaido University Hospital, Sapporo City General Hospital, Hokuyukai Neurological Hospital, Obihiro Kosei Hospital, Kushiro Rosai Hospital, Japanese Red Cross Asahikawa Hospital, Hakodate Municipal Hospital, and Wakkanai City Hospital) in Hokkaido, Japan from March 2012 to February 2013. MSA diagnoses were made in accordance with the second consensus statement including MRI findings [31]. The raters met each patient at 6-month intervals and scored them on the UMSARS, SARA, BBS, MSA-QoL, and SCOPA-AUT. Each evaluation of this study was performed on days when each patient's general condition was stable, and to the extent possible, at similar times of the day. The raters were seven board-certified neurologists, and a total of 217 examinations were performed. Each neurologist examined patients 3 to 163 times (one rater visited five hospitals over the course of 1 year). Patients' responses on each scale (UMSARS Part 1 = 12 items, UMSARS Part 2 = 14 items, SARA = 8 items, BBS = 14 items, MSA-QoL = 41 items, and SCOPA-AUT = 25 items) were also recorded. No interventions were conducted during the study period

except for those that patients had begun prior to study commencement. Amassed data were anonymized, and statistical analyses were performed.

This study was approved by the Institutional Review Board of Hokkaido University Hospital. Prior to starting this study, written informed consent was obtained from all participants. Patients with severe cognitive impairment were excluded.

Statistical Analysis

Statistical analysis was performed using JMP®Pro 10.0.0 (SAS Institute Inc., Cary, NC, USA). Score changes were evaluated using SRMs. SRMs were obtained by dividing the average score change by the standard deviation of score change. An SRM >0.8 was considered large, 0.5 to 0.8 moderate, and <0.5 small [32]. We interpreted a larger SRM as indicative of more rapid symptom progression. The association between score changes between each index and visit was evaluated using the Jonckheere-Terpstra trend test [33]. The correlations between disease duration and each score were also examined using Spearman's rank method. Stratified analysis was conducted for gender, symptom type, disease duration, and onset age. Sample size calculations required for interventional trials were conducted using the sample size formula for analysis of covariance with baseline variable adjustment similar to that of a previous study [18].

Results

We had initially recruited 87 patients, but excluded two because of inadequate diagnosis. The first assessment was conducted on 85 patients. Six months later, 69 patients were reevaluated, and 16 patients were unable to attend for various reasons. An additional 6 months later, we evaluated 63 patients. In the latter 6 months, three patients died and three patients could not attend (Fig. 1). The mean interval between the first and second assessment and the first and third assessment was 184.5 ± 13.1 and 363.8 ± 14.9 days, respectively.

Patient demographics are shown in Table 1. The proportion of patients was as follows: 29.4 % had MSA-P and 70.6 % had MSA cerebellar ataxia dominant subtype (MSA-C; 47 female and 38 male). The mean age at the first assessment was 63.8 years. The most frequent initial symptom was gait disturbances in MSA-C and MSA-P. Disturbances of fine dexterity movements of hands were significantly more frequent in MSA-P than in MSA-C. All MSA-P patients in this study had Parkinsonism onset. With regard to medication at baseline, l-dopa was most prescribed in MSA-P and taltirelin in MSA-C. At the first assessment, the scores of the UMSARS Part 1 and 2 were significantly higher among MSA-P. The majority of patients had been diagnosed with MSA 2–3 years earlier and was between 60 and 69 years of

age at the time of diagnosis (Fig. 2).

The comparison of total raw scores for each scale is shown in Fig. 3. For reference, the Barthel index [34] is also shown. UMSARS (parts 1, 2, and 4), SARA, and MSA-QoL scores increased, while BBS and Barthel index scores decreased over time. This score pattern reflects the typical course of MSA. UMSARS (parts 2 and 4), SARA, BBS, MSA-QoL, and Barthel index scores changed significantly over 12 months, but the Wilcoxon's rank method and the Jonckheere-Terpstra trend test could not reveal which scale had the greatest score changes. UMSARS (parts 1, 2, and 4), SARA, BBS, MSA-QoL, and Barthel index scores were significantly correlated with disease duration (Table 2). These scales were also significantly correlated with each other.

The SRMs of total scores are shown in Fig. 4a. The SRM of the UMSARS Part 4 was largest at 6 months, and that of the SARA was largest at 12 months. After limiting the analysis to only patients with good ADL (UMSARS Part $4 \leq 3$; i.e., ambulatory patients), the SRMs of the UMSARS (parts 2 and 4), SARA, and BBS were also larger, and the SRM of the BBS increased further. However, among patients with poor ADL, the SRM of UMSARS Part 4 was small, and that of SARA and UMSARS Part 2 were large; the SRM of UMSARS Part 4 did not change considerably in such cases (Table 3). The SRM of BBS and onset age were significantly correlated (Spearman's rank method, $\rho = 0.26$, $p = 0.03$).

Stratified analyses indicated a similar tendency. Figure 4b, c, and d shows the larger SRM scales. The SRMs of each scale did not differ between males and females. The SRMs were relatively high for MSA-P,

patients with disease duration of less than 4 years, and patients with an onset age of less than 62 years. An analysis comparing probable MSA and possible MSA is not reported because most of the patients in this study were probable MSA cases (Table 1).

Detailed analyses of individual items revealed that the SRMs of items such as skilled hand activities, walking, and standing were large. There was a similar trend in patients with a better ADL score. We composed a provisional scale including eight items with the largest SRMs (Table 4; SARA-1, gait; BBS-5, transfers; UMSARS Part 2–8, finger tapping; UMSARS Part 1–3, handwriting; BBS-7, standing with feet together; BBS-10, turning trunk [feet fixed]; BBS-11, turning 360°; and UMSARS Part 2–13, body sway) where the total SRM of this scale was 0.990 at 6 months and 1.285 at 12 months (Fig. 4e). This effect was more pronounced in patients with better ADL. Further, the total SRM of patients with UMSARS Part 4 score of 3 or less within this provisional scale was 0.976 at 6 months and 1.372 at 12 months. Not only did it take 5 min for a complete assessment, but this provisional scale correlated with the many other scales adopted in our study except for the UMSARS Part 3 and SCOPA-AUT. We performed sample size calculations required for interventional trials. Ninety-eight patients would be needed for a 30 % improvement under an 80 % statistical power for this provisional scale. This number was smaller than those of other scales in our investigation, including the UMSARS Part 2 (Fig. 5 and Table 5).

Discussion

This is the first study that directly compares MSA scales while including SRM values. The follow-up periods were 6 to 12 months because we were aware that clinical trials place emphasis on score changes over a short period of time. SRM is one of the indicators of responsiveness, and the advantage is that scales with different total scores can be compared. Additionally, a large SRM indicates rapid disease progression, and a decrease in SRM reflects an effect of treatment. Thus, to facilitate the development of treatment for MSA, the application of SRMs in clinical trials is advantageous.

The patients in this study were a good representation of the Asian and Japanese MSA population and were consistent with the previous studies [35–37]. Our study patients were in the relatively early stages of MSA and consequently had good ADL, but the average onset age was slightly high. Moreover, they had received the standard available treatments.

When we compared the raw scores of each scale, the differences were unclear; however, the SRMs provided clarification. The SRM of the UMSARS Part 4 was largest at 6 months, and that of SARA was largest at 12 months. Since the SRMs of UMSARS Part 2 and BBS were also relatively large, these scales were assumed to be sensitive to short-term changes in MSA symptoms. The UMSARS, used in much clinical research on MSA, showed relatively high responsiveness. This complements available data suggesting the usefulness of UMSARS, although the scale included items without obvious score changes. A medical interview, such as UMSARS Part 1 and self-completed questionnaires like the MSA-QoL and SCOPA-AUT, is time efficient; they are helpful in assessing subjective symptoms [38], but are unable to

adequately capture changes in symptoms because their SRMs are relatively low.

The SRMs for MSA-P (The initial symptom of All the MSA-P patients in this study was parkinsonism), the shorter duration group, and the early onset group were larger than those for MSA-C, the longer duration group, and the late onset group. The former groups were graded to have a rapid progression. The poor prognosis of MSA-P and rapid progression of early-stage MSA patients were previously mentioned [17, 35], but it was not indicated that early-onset MSA would rapidly progress.

The present results reconfirm that the recruitment of ambulatory patients is essential for clinical trials. Moreover, we suggest that walking and standing may indicate better ADL in MSA patients. In contrast, the content affecting ADL, such as that assessed in the UMSARS Part 4, shows a ceiling effect, meaning that the skilled hand activities of the UMSARS Part 2 or SARA are points of focus for patients with worse ADL.

Overall scores on scales other than the UMSARS Part 3 and Part 4, which only contain one item each, could have potentially been influenced by differences in individual item scores. Thus, in those cases, score changes were cancelled out. Detailed analyses showed that the SRMs of individual items such as skilled hand activities, walking, and standing were large. These items are thought to be rapidly changing symptoms in the MSA trajectory. These are also influenced by cerebellar ataxia and Parkinsonism. Gait ataxia is thought to derive from deficits in the inferior olivary nucleus, the pontine nuclei, and the Purkinje cell layer of the cerebellum. In particular, reflecting rapid symptom changes of walking and

standing, atrophy of vermis might progress faster. Retropulsion or bradykinesia arises from nigrostriatal impairment, and skilled hand activities result from both olivopontocerebellar and nigrostriatal impairment. This pathogenesis was consistent with the pathological changes in MSA.

Comprehensive scales assessing ataxia, Parkinsonism, and autonomic dysfunction are important, whereas tools that are sensitive to symptom changes are also useful. A provisional scale composed of the eight items with the largest SRMs (gait, transfers, finger tapping, handwriting, standing with feet together, turning trunk (feet fixed), turning 360°, and body sway) was considered, and the total SRM of this scale over 6 and 12 months was found to be larger than all over scales included in this study. In addition, this scale not only required a far shorter amount of time for the examination compared to UMSARS (~5 as opposed to 20 min), but the required sample size was smaller. This provisional scale which includes these eight items might have with the potential to sensitively detect changes in MSA symptoms that occur over a short period. Clinical trials for MSA, such as those that examined the effects of minocycline [39] and rifampicin [40], have not yet yielded an effective treatment for MSA. This may have been due to inadequate sample size, study design, and outcome measures. However, the use of a more sensitive scale may yield more promising results. Slowing disease progression and improving skilled hand activities, walking, and standing will improve ADL. We think this provisional scale, which can evaluate these symptoms, could be beneficial and suitable for early-stage MSA patients. Moreover, as this scale was correlated with many other scales, i.e., this provisional scale had criterion-related validity, further

incorporation of this scale will greatly reduce overall assessment time. However, the content validity and the reliability of these eight items should be assessed in the future.

We think symptom assessment scales focused on gait and balance disturbances are instrumental for the evaluation of patients in early stages of MSA. Furthermore, the Spinocerebellar Ataxia Functional Index (including 8 m timed walk, PATA rate, and 9-hole pegboard test) and the Functional Independent Measure are also useful for evaluating ADL [41, 42], but these measurements need several items of equipment and considerable time. In light of the simple assessments preferred in outpatient departments, we did not adopt these measurements nor the NNIPPS scale.

The SCOPA-AUT does not adequately assess urinary disturbances when such disturbances progress, and patients need intermittent catheterization or transurethral retained catheterization. In those cases, the SCOPA-AUT score decreases. In addition, the SCOPA-AUT total score may underestimate autonomic dysfunction. Thus, the progression of autonomic dysfunction may not be evaluated precisely. Autonomic dysfunction is a very important factor in MSA progression [17, 43], but these SRMs were small. Existing scales could not capture accurate changes in autonomic dysfunction. It is consequently necessary to review various measurement methods, including objective neurophysiological examinations or neuroimaging techniques. Further, the existing MSA scales did not attach high values to vocal cord dysfunction or sleep apnea, which can cause sudden death. Such factors should be considered in the revision of comprehensive scales such as UMSARS.

The limitations of this study were as follows: (i) while patients had received a clinical diagnosis of MSA, this was not pathologically confirmed; (ii) since it was difficult to match onset ages and disease durations, the time point for estimating scale scores or SRMs varied; (iii) while the present results can be compared with those of studies conducted in Western countries, such a comparison is not perfect owing to the smaller number of MSA-P patients in the present sample; (iv) this study mainly dealt with semi-quantitative assessments, and therefore quantitative tests for cerebellar ataxia based on neurophysiological tasks should be considered as a future extension; and (v) clinical symptom scales showed a ceiling effect. This ceiling effect could potentially be reduced by including study patients with better ADL. However, a measurement method that changes linearly over the course of all disease stages has not been developed.

Conclusion

The UMSARS (parts 2 and 4), SARA, and BBS were identified as probable sensitive scales for assessing changes in MSA symptoms over 12 months. Among MSA-P patients, the short duration group and early onset group showed rapid progression. Detailed analyses revealed that items with large SRMs, such as skilled hand activities, walking, and standing, were thought to be useful as a way to evaluate short-term changes.

Acknowledgements

We thank all patients and doctors* for their active cooperation.

*Shigehisa Ura, Kiyomi Kuroshima (Japanese Red Cross Asahikawa Hospital), Shinsuke Hamada (Hokuyukai Neurological Hospital), Kazuhiro Horiuchi (Obihiro Kosei Hospital), Shinichi Shirai (Kushiro Rosai Hospital), Takahiro Kano, Makoto Hirotani, Ikuko Takahashi, Hisashi Uwatoko, and Mari Kimura (Hokkaido University Hospital).

References

[1] Graham JG, Oppenheimer DR. Orthostatic hypotension and nicotine sensitivity in a case of multiple system atrophy. JNNP 1969; 32: 28-34

[2] Quinn N. Multiple system atrophy--the nature of the beast. JNNP 1989; Suppl: 78-89

[3] Wakabayashi K, Yoshimoto M, Tsuji S, Takahashi H. Alpha-synuclein immunoreactivity in glial cytoplasmic inclusions in multiple system atrophy. Neurosci Lett 1998; 249: 180-182

[4] Orosz F, Kovács GG, Lehotzky A, Oláh J, Vincze O, Ovádi J. TPPP/p25: from unfolded protein to misfolding disease: prediction and experiments. Biol Cell 2004; 96: 701-711

[5] Kaufman E, Hall S, Surova Y, Widner H, Hansson O, Lindqvist D. Proinflammatory cytokines are elevated in serum of patients with multiple system atrophy. PLoS One 2013; 8: e62354

[6] Multiple-System Atrophy Research Collaboration. Mutations in COQ2 in familial and sporadic multiple-system atrophy. N Engl J Med 2013; 369: 233-244

[7] Ahmed Z, Asi YT, Sailer A, et al. The neuropathology, pathophysiology and genetics of multiple system atrophy. Neuropathol Appl Neurobiol 2012; 38: 4-24

[8] Ozawa T, Sekiya K, Sekine Y, et al. Maintaining glottic opening in multiple system atrophy: efficacy of serotonergic therapy. Mov Disord 2012; 27: 919-921

[9] Novak P, Williams A, Ravin P, Zurkiya O, Abduljalil A, Novak V. Treatment of multiple system atrophy using intravenous immunoglobulin. BMC Neurol 2012; 12: 131

[10] Lee PH, Lee JE, Kim HS, et al. A randomized trial of mesenchymal stem cells in multiple system

atrophy. Ann Neurol 2012; 72: 32-40

[11] Luo Y, Fan Y, Zhou B, Xu Z, Chen Y, Sun X. Generation of induced pluripotent stem cells from skin fibroblasts of a patient with olivopontocerebellar atrophy. Tohoku J Exp Med 2012; 226: 151-159

[12] 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

[13] Fahn S, Elton RL, members of the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. In: Fahn D, Marsden CD, Calne D, Goldstein M, editors. Recent Development in Parkinson's Disease, vol. 2. Florham Park, NJ: Maacmillan Healthcare Information. 1987; p153-163, 293-304

[14] Trouillas P, Takayanagi T, Hallett M, et al. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropharmacology Committee of the World Federation of Neurology. J Neurol Sci 1997; 145: 205-211

[15] Suarez GA, Opfer-Gehrking TL, Offord KP, Atkinson EJ, O'Brien PC, Low PA. The Autonomic Symptom Profile: a new instrument to assess autonomic symptoms. Neurology 1999; 52: 523-528

[16] Geser F, Seppi K, Stampfer-Kountchev M, et al. The European Multiple System Atrophy-Study Group (EMSA-SG). J Neural Transm 2005; 112: 1677-1686

[17] Wenning GK, Geser F, Krismer F, et al. The natural history of multiple system atrophy: a prospective European cohort study. Lancet Neurol 2013; 12: 264-274

[18] May S, Gilman S, Sowell BB, et al. Potential outcome measures and trial design issues for multiple system atrophy. Mov Disord 2007; 22: 2371-2377

[19] Schmitz-Hübsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 2006; 66: 1717-1720

[20] Yabe I, Matsushima M, Soma H, Basri R, Sasaki H. Usefulness of the Scale for Assessment and Rating of Ataxia (SARA). J Neurol Sci 2008; 266: 164-166

[21] Berg K, Wood-Dauphinée S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. Physiother Can 1989; 41: 304-311

[22] Qutubuddin AA, Pegg PO, Cifu DX, Brown R, McNamee S, Carne W. Validating the Berg Balance Scale for patients with Parkinson's disease: a key to rehabilitation evaluation. Arch Phys Med Rehabil 2005; 86: 789-792

[23] Oz G, Iltis I, Hutter D, Thomas W, Bushara KO, Gomez CM. Distinct neurochemical profiles of spinocerebellar ataxias 1, 2, 6, and cerebellar multiple system atrophy. Cerebellum 2011; 10: 208-217

[24] Matsushima M, Yabe I, Uwatoko H, Shirai S, Hirotani M, Sasaki H. Reliability of the Japanese version of the Berg balance scale. Intern Med 2014; 53: 1621-1624

[25] Schrag A, Selai C, Mathias C, et al. Measuring health-related quality of life in MSA: the MSA-QoL. Mov Disord 2007; 22: 2332-2338

[26] Meissner WG, Foubert-Samier A, Dupouy S, et al. Assessment of quality of life with the multiple

system atrophy health-related quality of life scale. Mov Disord 2012; 27: 1574-1577

[27] Visser M, Marinus J, Stiggelbout AM, Van Hilten JJ. Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. Mov Disord 2004; 19: 1306-1312

[28] Damon-Perrière N, Foubert-Samier A, De Cock VC, et al. Assessment of the Scopa-Aut questionnaire in multiple system atrophy: relation to UMSARS scores and progression over time. Parkinsonism Relat Disord 2012; 18: 612-615

[29] Payan CA, Viallet F, Landwehrmeyer BG, et al. Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: validation of the NNIPPS--Parkinson Plus Scale. PLoS One 2011; 6: e22293

[30] Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990; 28: 632-642

[31] Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology 2008; 71: 670-676

[32] Cohen J. Statistical power analysis for the behavioral sciences 2nd edition. 1988; p20-26

[33] Jonckheere AR. A distribution-free k-sample test against ordered alternatives. Biometrika 1954; 41: 133-145

[34] Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965; 14: 61-65

[35] 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

[36] Yabe I, Soma H, Takei A, Fujiki N, Yanagihara T, Sasaki H. MSA-C is the predominant clinical phenotype of MSA in Japan: analysis of 142 patients with probable MSA. J Neurol Sci 2006; 249: 115-121

[37] Jamora RD, Gupta A, Tan AK, Tan LC. Clinical characteristics of patients with multiple system atrophy in Singapore. Ann Acad Med Singapore 2005; 34: 553-557

[38] Matsushima M, Yabe I, Hirotani M, Kano T, Sasaki H. Reliability of the Japanese version of the Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire. Clin Neurol Neurosurg 2014; 124: 182-184

[39] Dodel R, Spottke A, Gerhard A, et al. Minocycline 1-year therapy in multiple-system-atrophy: effect on clinical symptoms and [(11)C] (R)-PK11195 PET (MEMSA-trial). Mov Disord 2010; 25: 97-107

[40] Low PA, Robertson D, Gilman S, et al. Efficacy and safety of rifampicin for multiple system atrophy: a randomised, double-blind, placebo-controlled trial. Lancet Neurol 2014; 13: 268-275

[41] Schmitz-Hübsch T, Giunti P, Stephenson DA, et al. SCA Functional Index: a useful compound performance measure for spinocerebellar ataxia. Neurology 2008; 71: 486-492

[42] Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. Adv Clin Rehabil 1987; 1: 6-18

[43] Calandra-Buonaura G, Guaraldi P, Sambati L, et al. Multiple system atrophy with prolonged survival:

is late onset of dysautonomia the clue? Neurol Sci 2013; 34: 1875-1878

Legends

Table 1. Demographic data of the patients at the first evaluation

Total n = 85, * Fisher's exact test and Wilcoxon's rank test

Table 2. The correlation between disease duration and each scale score

ρ Spearman's rank coefficients, * $p < 0.05$

Table 3. Standardized response means (SRMs) of each scale for all patients, Unified Multiple System Atrophy Rating Scale (UMSARS) Part 4 ≤ 3, UMSARS Part 4 ≤ 2, and UMSARS Part 4 ≥ 4

SRM6: SRMs in six months, SRM12: SRMs in twelve months

Table 4. The eight items with the largest standardized response mean

Table 5. The required sample size to detect treatment effects for each scale

UMSARS Unified Multiple System Atrophy Rating Scale, SARA Scale for the Assessment and Rating of Ataxia, BBS Berg Balance Scale, MSA-QoL Multiple System Atrophy Health-Related Quality of Life scale, SCOPAyAUT Scales for Outcomes in Parkinson's Disease–Autonomic questionnaire

Figure 1. Study procedure

Figure 2. Patient distribution of onset age and disease duration

Figure 3. The comparison of raw scores revealed time-dependent worsening among most scales.

UMSARS1 Unified Multiple System Atrophy Rating Scale (UMSARS) Part 1, UMASRS2 UMSARS Part 2, UMSARS3Δs UMSARS Part 3 systolic decrease, UMSARS3Δd UMSARS Part 3 diastolic decrease, UMSARS4 UMSARS Part 4, SARA Scale for the Assessment and Rating of Ataxia, BBS Berg Balance Scale, MSA-QoL Multiple System Atrophy Health-Related Quality of Life scale, SCOPA-AUT Scales for Outcomes in Parkinson's Disease–Autonomic questionnaire.

Figure 4. Standardized response means (SRMs) of total scores of each scale.

a SRMs in all patients. b SRMs in MSA Parkinsonism-dominant subtype (MSA-P) and MSA cerebellar ataxia dominant subtype (MSA-C). c SRMs in the groups of disease duration less than 4 years and over 4 years. d SRMs in patients with an onset age of less than 62 years and greater than 62 years. e SRMs of a pilot scale consisting of the largest SRM items. SRM6 SRM in 6 months, SRM12 SRM in 12 months

Figure 5. Sample size calculations.

Required number (per group) of participants to measure treatment effect with 80 and 90 % statistical power for the provisional scale which consists of selected eight items with the largest standardized response mean (left) and Unified Multiple System Atrophy Rating Scale (UMSARS) Part 2 (right)

Table 1. Demographic data of the patients at the first evaluation

	Total	MSA-P	MSA-C	p value*
Patients, n (probable MSA, n)	85 (82)	25 (25)	60 (57)	0.5520
Gender, female (%)	47 (55)	11 (44)	36 (60)	0.2325
Onset age, years, mean ± SD	60.0 ± 8.3	60.8 ± 6.7	59.7 ± 9.0	0.9192
Disease duration, months, mean ± SD	45.5 ± 29.3	52.3 ± 32.6	42.7 ± 27.8	0.1994
Initial symptoms (include overlap)				
Dysarthria	9	2	7	0.6251
Upper extremity disturbances	13	7	6	0.0375
Gait disturbances	65	14	51	0.0044
Autonomic disturbances	6	3	3	0.2583
Treatment (include overlap)				
L-dopa	24	19	5	< 0.0001
Taltirelin	61	14	47	0.0389
Vasopressors	9	2	7	0.6641
Gastrostomy, n	3	1	2	0.8664
Tracheostomy, n	2	1	1	0.4995
Scale scores, mean ± SD				
UMSARS Part 1	20.4 ± 9.5	24.0 ± 8.6	17.8 ± 8.6	0.0197
UMSARS Part 2	20.9 ± 9.6	24.4 ± 9.0	18.1 ± 8.3	0.0208
UMSARS Part 3 (systolic decrease)	18.5 ± 16.1	17.5 ± 14.8	18.1 ± 18.1	0.7212
UMSARS Part 3 (diastolic decrease)	6.1 ± 11.0	7.4 ± 12.2	4.2 ± 10.6	0.4292
UMSARS Part 4	2.9 ± 1.2	2.9 ± 1.0	2.7 ± 1.1	0.6793
SARA	18.4 ± 7.2	17.0 ± 7.5	17.9 ± 6.7	0.4430
BBS	26.9 ± 17.3	27.5 ± 15.5	28.3 ± 17.1	0.9270
MSA-QoL	56.2 ± 33.6	56.4 ± 29.8	51.7 ± 35.8	0.3157
SCOPA-AUT	14.6 ± 7.4	17.5 ± 9.9	13.3 ± 6.2	0.1323
Barthel index	68.9 ± 28.3	62.9 ± 27.5	75.7 ± 24.6	0.1040

UMSARS, Unified Multiple System Atrophy Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia; BBS, Berg Balance Scale; MSA-QoL, Multiple System Atrophy Health-Related Quality of Life scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease – Autonomic questionnaire.

Table 2. Correlations between disease duration and each scale score

	First assessment	Second assessment	Third assessment
UMSARS Part 1	ρ = 0.4883 p < 0.0001*	ρ = 0.3245 p = 0.0065*	ρ = 0.3406 p = 0.0063*
UMSARS Part 2	ρ = 0.5067 p < 0.0001*	ρ = 0.3817 p = 0.0012*	ρ = 0.3408 p = 0.0063*
UMSARS Part 3 (systolic decrease)	ρ = 0.0517 p = 0.6447	ρ = -0.0232 p = 0.8555	ρ = -0.0162 p = 0.9057
UMSARS Part 3 (diastolic decrease)	ρ = 0.0707 p = 0.5278	ρ = 0.0855 p = 0.5015	ρ = 0.1183 p = 0.3850
UMSARS Part 4	ρ = 0.4216 p < 0.0001*	ρ = 0.3275 p = 0.0060*	ρ = 0.3901 p = 0.0016*
SARA	ρ = 0.4279 p < 0.0001*	ρ = 0.3486 p = 0.0033*	ρ = 0.3242 p = 0.0095*
BBS	ρ = -0.5616 p < 0.0001*	ρ = -0.4991 p < 0.0001*	ρ = -0.4299 p = 0.0004*
MSA-QoL	ρ = 0.3112 p = 0.0037*	ρ = 0.2237 p = 0.0647	ρ = 0.2037 p = 0.1094
SCOPA-AUT	ρ = -0.0179 p = 0.8715	ρ = -0.0915 p = 0.4546	ρ = -0.0833 p = 0.5164
Barthel index	ρ = -0.5285 p < 0.0001*	ρ = -0.3958 p = 0.0021*	ρ = -0.4145 p = 0.0015*

UMSARS, Unified Multiple System Atrophy Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia; BBS, Berg Balance Scale; MSA-QoL, Multiple System Atrophy Health-Related Quality of Life scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease – Autonomic questionnaire.

Table 3. SRMs of each scale for all patients, Unified Multiple System Atrophy Rating Scale (UMSARS) Part 4 ≤ 3, UMSARS Part 4 ≤ 2 and UMSARS Part 4 ≥ 4

n	SRM6	UMSARS Part 1	UMSARS Part 2	UMSARS Part 3 (systolic decrease)	UMSARS Part 3 (diastolic decrease)	UMSARS Part 4	SARA	BBS	MSA-QoL	SCOPA-AUT
69	All patients	0.218	0.557	0.133	0.155	0.796	0.702	0.688	0.309	-0.082
49	UMSARS Part 4 ≤ 3	0.171	0.551	0.182	0.120	1.120	0.677	0.699	0.336	-0.158
29	UMSARS Part 4 ≤ 2	0.510	0.437	0.252	0.134	0.800	0.762	0.880	0.402	-0.113
20	UMSARS Part 4 ≥ 4	0.301	0.586	-0.057	0.358	0.181	0.746	0.643	0.245	0.092

n	SRM12	UMSARS Part 1	UMSARS Part 2	UMSARS Part3 (systolic decrease)	UMSARS Part 3 (diastolic decrease)	UMSARS Part 4	SARA	BBS	MSA-QoL	SCOPA-AUT
63	All patients	0.419	1.092	0.046	0.112	0.863	1.096	1.074	0.577	-0.044
46	UMSARS Part 4 ≤ 3	0.380	1.073	0.039	-0.037	1.125	1.043	1.199	0.577	-0.090
26	UMSARS Part 4 ≤ 2	0.462	0.971	0.047	-0.148	0.875	1.081	1.361	0.536	-0.039
17	UMSARS Part 4 ≥ 4	0.511	1.117	0.060	0.451	0.313	1.259	0.802	0.596	0.104

SRM	
Largest	
Second	
Third	
Fourth	

UMSARS, Unified Multiple System Atrophy Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia; BBS, Berg Balance Scale; MSA-QoL, Multiple System Atrophy Health-Related Quality of Life scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease – Autonomic questionnaire.

Table 4. The eight items with the largest standardized response mean

1	Gait (from SARA 1)
	Proband is asked (1) to walk at a safe distance parallel to a wall including a half-turn (turn around to face the opposite direction of gait) and (2) to walk in tandem (heel-to-toe) without support. 0. Normal, no difficulties in walking, turning, or walking tandem (up to one misstep allowed) 1. Slight difficulties, only visible when walking 10 consecutive steps in tandem 2. Clearly abnormal, tandem walking >10 steps not possible 3. Considerable staggering, difficulties in half-turn, but without support 4. Marked staggering, intermittent support of the wall required 5. Severe staggering, permanent support of one stick or light support by one arm required 6. Walking > 10 m only with strong support (two special sticks or stroller or accompanying person) 7. Walking < 10 m only with strong support (two special sticks or stroller or accompanying person) 8. Unable to walk, even if supported
2	Transfers (from BBS 5)
	Arrange chairs(s) for a pivot transfer. Ask subject to transfer one way toward a seat with armrests and one way toward a seat without armrests. Two chairs (one with and one without armrests) or a bed and a chair may be used. 0. Able to transfer safely with minor use of hands 1. Able to transfer safely definite need of hands 2. Able to transfer with verbal cueing and/or supervision 3. Needs one person to assist 4. Needs two people to assist or supervise to be safe
3	Finger tapping (from UMSARS Part 2-8)
	Patient taps thumb with index finger in rapid succession with widest amplitude possible, with each hand for at least 15 to 20 seconds. Rate the worst affected limb. Note that impaired performance on this task can be caused by bradykinesia and/or cerebellar incoordination. Rate functional performance regardless of underlying motor disorder. 0. Normal. 1. Mildly impaired. 2. Moderately impaired. 3. Severely impaired. 4. Can barely perform the task.
4	Handwriting (from UMSARS Part 1-3)
	0. Normal 1. Mildly impaired (all words are legible). 2. Moderately impaired (up to half of the words are illegible). 3. Markedly impaired (the majority of words are illegible). 4. Unable to write.

5	Standing unsupported with feet together (from BBS 7)
	Place your feet together and stand without holding 0. able to place feet together independently and stand 1 minute safely 1. able to place feet together independently and stand for 1 minute with supervision 2. able to place feet together independently but unable to hold for 30 seconds 3. needs help to attain position but able to stand 15 seconds feet together 4. needs help to attain position and unable to hold for 15 seconds
6	Turning to look behind over left and right shoulders while standing (from BBS 10)
	Turn to look directly behind you over toward left shoulder. Repeat to the right. Examiner may pick an object to look at directly behind the subject to encourage a better twist turn. 0. looks behind from both sides and weight shifts well 1. looks behind one side only other side shows less weight shift 2. turns sideways only but maintains balance 3. needs supervision when turning 4. needs assist to keep from losing balance or falling
7	Turning 360 degrees (from BBS 11)
	Turn completely around in a full circle. Pause. Then turn a full circle in the other direction. 0. able to turn 360 degrees safely in 4 seconds or less 1. able to turn 360 degrees safely one side only in 4 seconds or less 2. able to turn 360 degrees safely but slowly 3. needs close supervision or verbal cueing 4. needs assistance while turning
8	Body sway (from UMSARS part 2-13)
	Rate spontaneous body sway and response to sudden, strong posterior displacement produced by pull on shoulder while patient erect with eyes open and feet slightly apart. Patient has to be warned. 0. Normal. 1. Slight body sway and/or retropulsion with unaided recovery. 2. Moderate body sway and/or deficient postural response; might fall if not caught by examiner. 3. Severe body sway. Very unstable. Tends to lose balance spontaneously. 4. Unable to stand without assistance.

UMSARS, Unified Multiple System Atrophy Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia; BBS, Berg Balance Scale

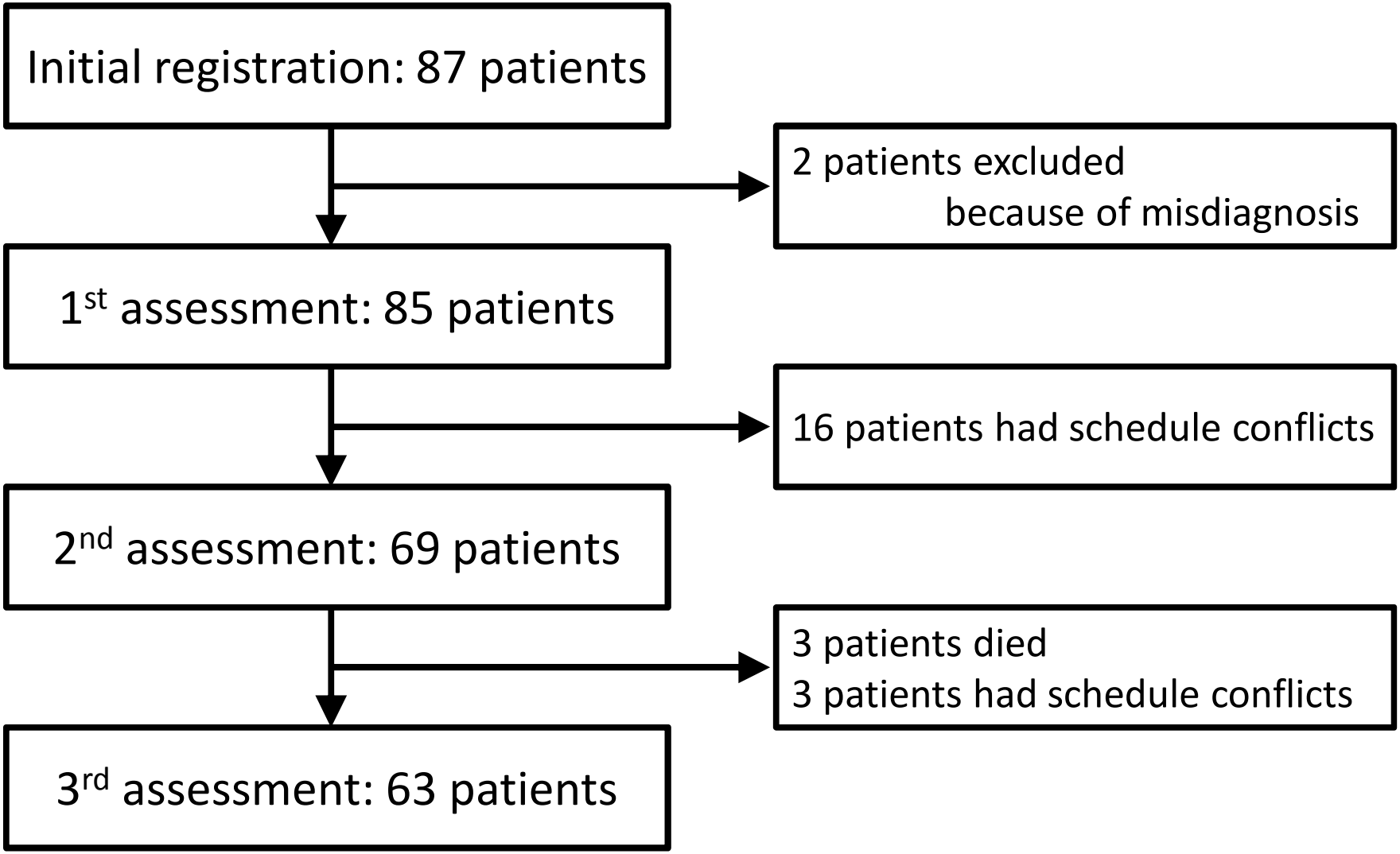
Table 5. The required sample size to detect treatment effects for each scale

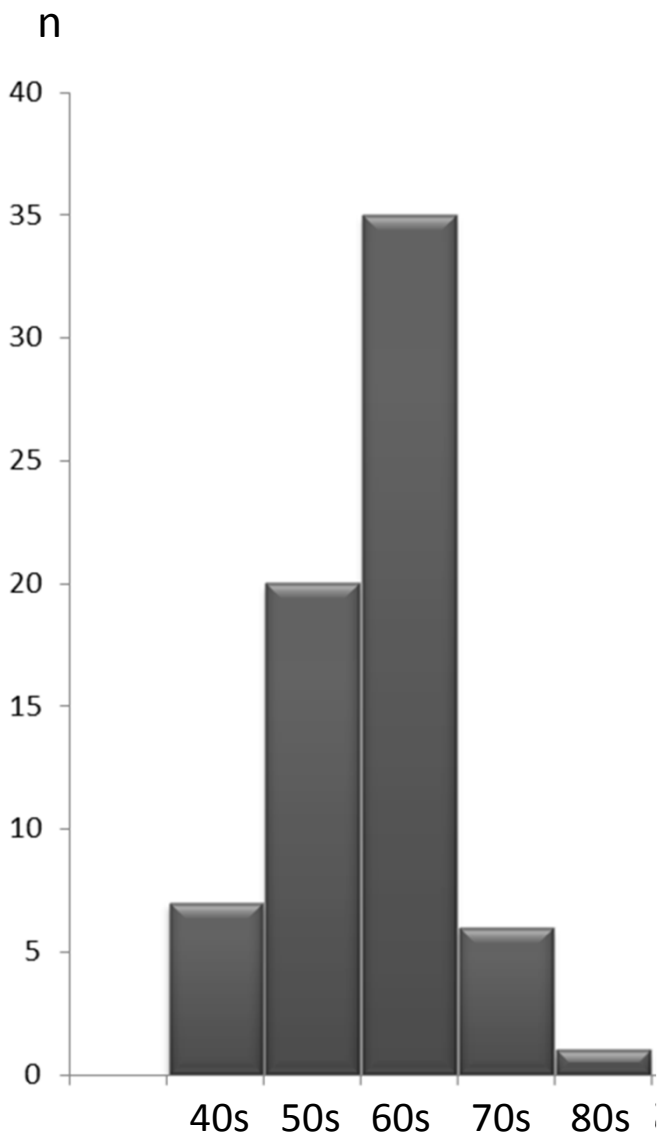
Scales	Sample size (n)*
UMSARS Part 1	1,002
UMSARS Part 2	140
UMSARS Part 3 (systolic decrease)	152,590
UMSARS Part 3 (diastolic decrease)	13,900
UMSARS Part 4	217
SARA	136
BBS	142
MSA-QoL	532
SCOPA-AUT	499,207
Barthel index	249
the provisional scale**	98

* In the case of 80% power, 30% treatment effect.

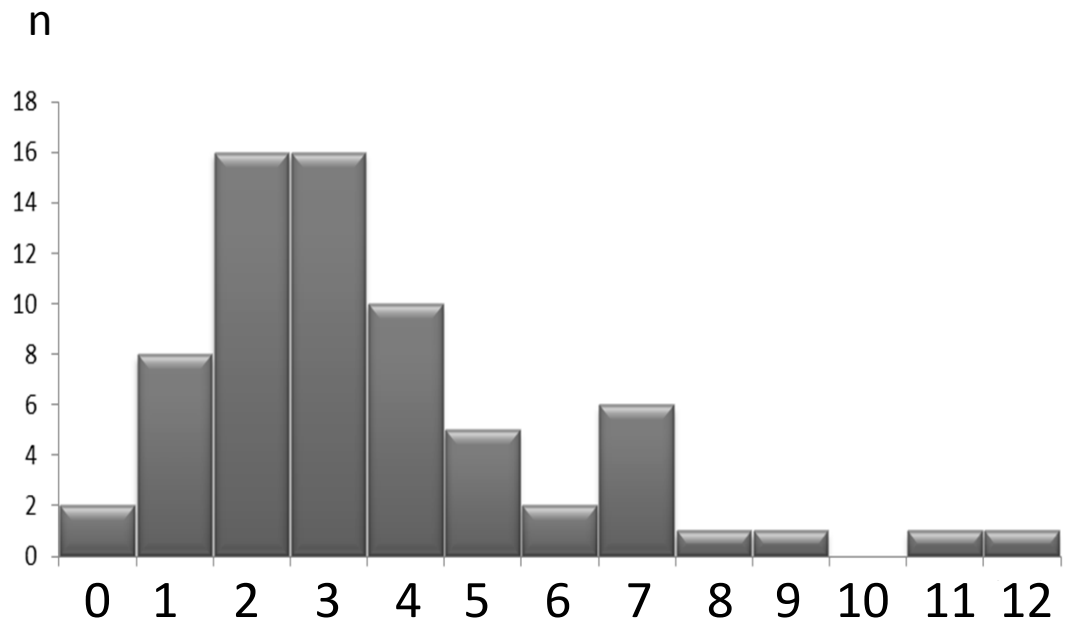
** The provisional scale consists of selected eight items with the largest standardized response mean.

UMSARS, Unified Multiple System Atrophy Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia; BBS, Berg Balance Scale; MSA-QoL, Multiple System Atrophy Health-Related Quality of Life scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease – Autonomic questionnaire; SiMSAS, Simple Multiple System Atrophy Scale.


Initial registration: 87 patients
2 patients excluded because of misdiagnosis
1st assessment: 85 patients
16 patients had schedule conflicts
2nd assessment: 69 patients
3 patients died
3 patients had schedule conflicts
3rd assessment: 63 patients

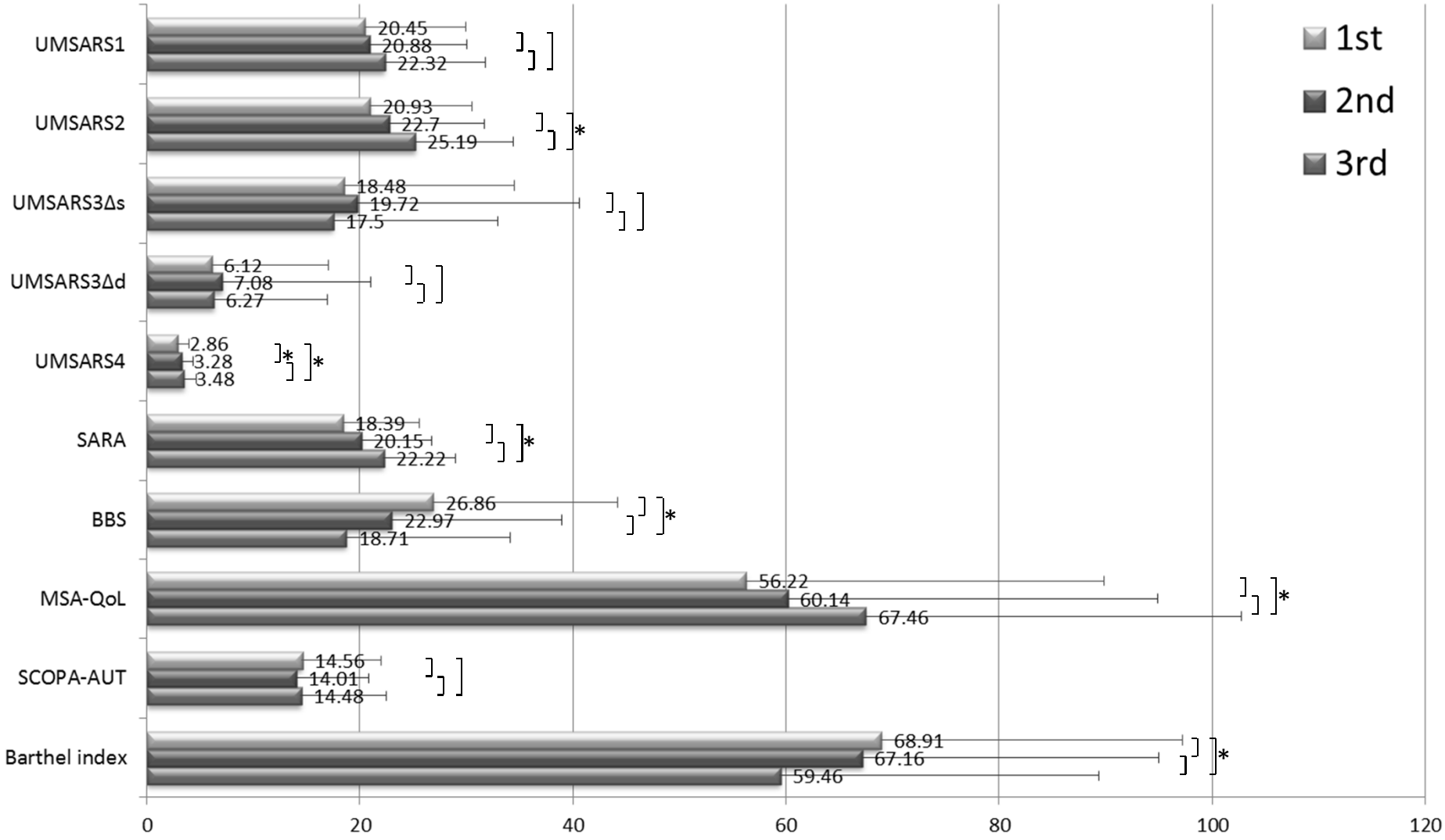


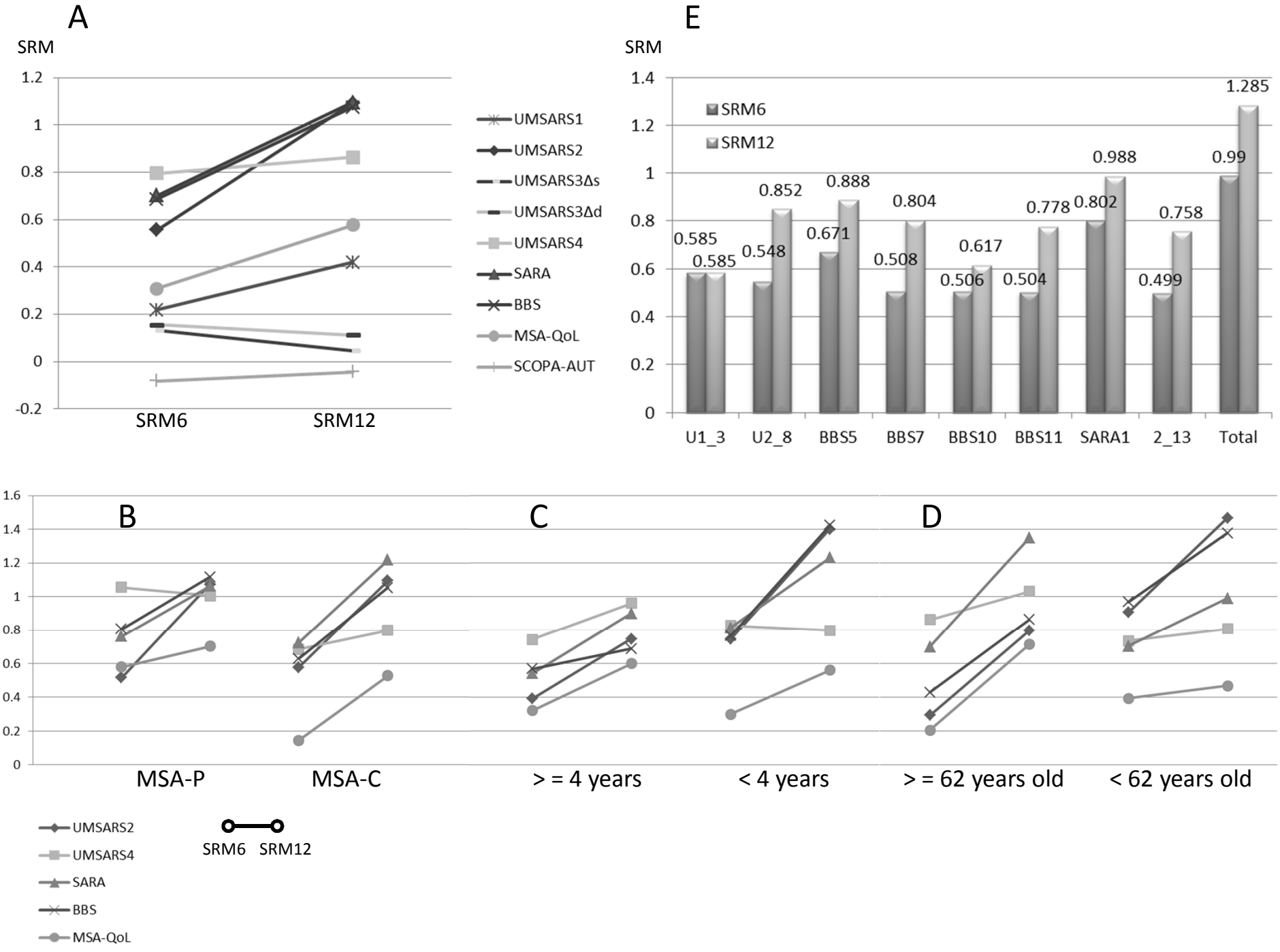
Onset age

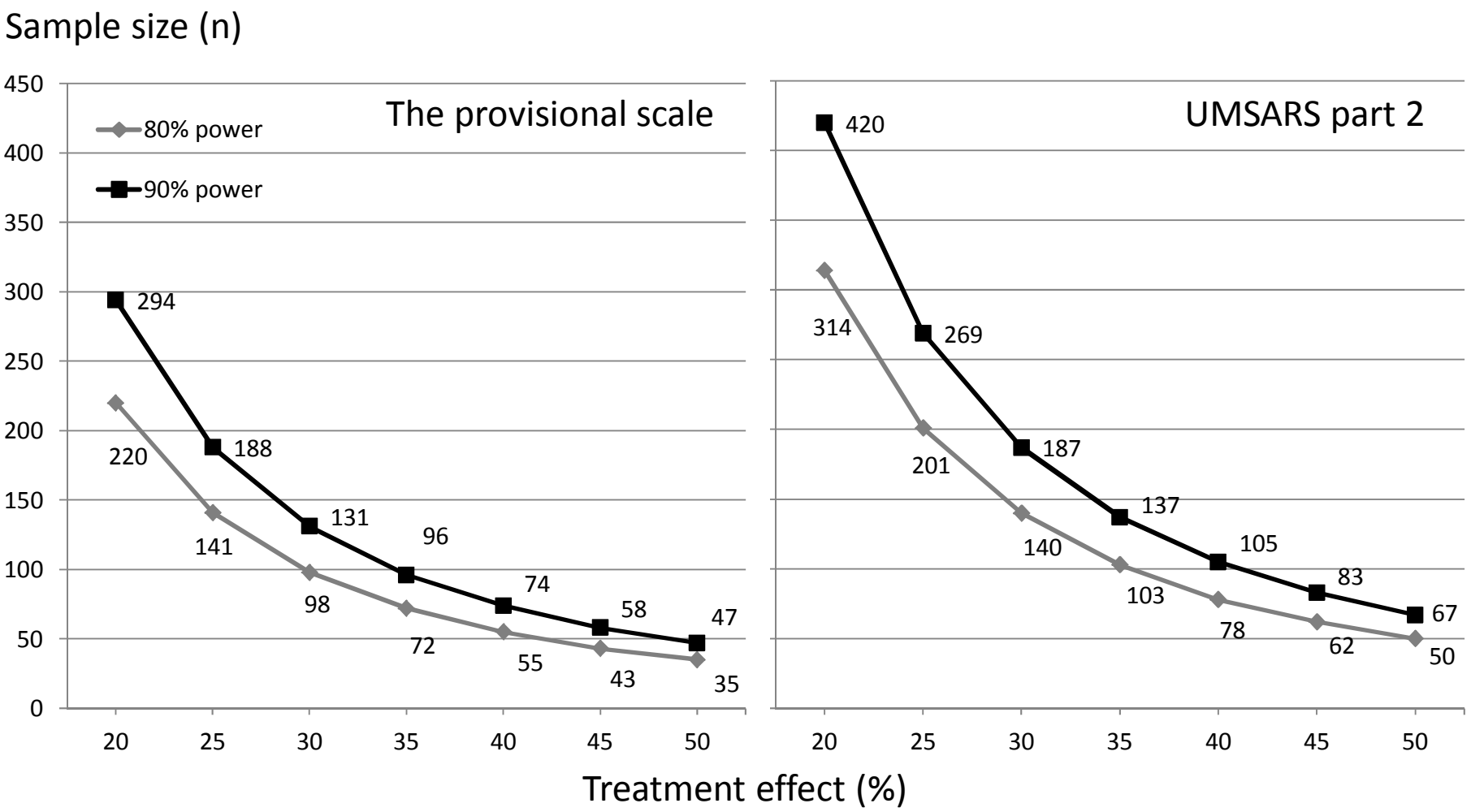


Disease duration (years)

* p < 0.05, Wilcoxon‘s rank test and Jonckheere-Terpstra trend test




A
SRM
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
SRM6
SRM12
UMSARS1
UMSARS2
UMSARS3Δs
UMSARS3Δd
UMSARS4
SARA
BBS
MSA-QoL
SCOPA-AUT
E
SRM
1.4
1.2
1
0.8
0.6
0.4
0.2
0
SRM6
SRM12
0.585
0.585
0.548
0.852
0.671
0.888
0.508
0.804
0.506
0.617
0.504
0.778
0.802
0.988
0.499
0.758
0.99
1.285
U1_3
U2_8
BBS5
BBS7
BBS10
BBS11
SARA1
2_13
Total
B
C
D
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
MSA-P
MSA-C
> = 4 years
< 4 years
> = 62 years old
< 62 years old
UMSARS2
UMSARS4
SARA
BBS
MSA-QoL
SRM6
SRM12


Sample size (n)
450
400
350
300
250
200
150
100
50
0
80% power
90% power
The provisional scale
294
188
131
96
74
58
47
220
141
98
72
55
43
35
UMSARS part 2
420
269
187
137
105
83
67
314
201
140
103
78
62
50
20
25
30
35
40
45
50
Treatment effect (%)