Axial motor clues to identify atypical parkinsonism: A multicentre European cohort study

Axial motor clues to identify atypical parkinsonism: A multicentre European cohort study

Carlijn D.J.M. Borm MD[1], Florian Krismer MD PhD[2], Gregor K. Wenning MD PhD[2], Klaus Seppi MD[2], Werner Poewe MD[2] ,Maria Teresa Pellecchia MD PhD,[3] Paolo Barone MD[3] PhD, Erik Lisbjerg Johnsen MD, PhD[4], Karen Østergaard MD PhD[4], Tanya Gurevich MD[5], Ruth Djaldetti MD[6], Luisa Sambati MD,PhD[7] , Pietro Cortelli MD PhD[7], Igor Petrović MD[8], Vladimir s. Kostić MD PhD[8], Hana Brožová MD PhD[9], Evžen Růžička MD[9], Maria Jose Marti MD PhD[10], Eduardo Tolosa MD PhD[10], Margherita Canesi MD PhD[11,] Bart Post MD PhD[1], Jorik Nonnekes MD PhD[12], Bastiaan R. Bloem MD PhD[1], on behalf of the European MSA Study Group (EMSA-SG)

[1]Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Neurology, Parkinson Centre Nijmegen (ParC) Nijmegen, The Netherlands
[2]Department of Neurology, Medical University Innsbruck, Austria
[3] Centre for Neurodegenerative Diseases, Department of Medicine and Surgery, University of Salerno, Italy
[4]Department of Neurology, University Hospital, Aarhus, Denmark
[5]Neurological Institute, Tel-Aviv Sourasky Medical Centre, Tel Aviv, Israel
[6]Department of Neurology, Rabin Medical Centre, Petach-Tiqva, Israel
[7]IRCCS Institute of Neurological Sciences of Bologna and Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
[8]Institute of Neurology, Clinical Centre of Serbia, Belgrade, Serbia; Faculty of Medicine, University of Belgrade, Belgrade, Serbia.
[9]Department of Neurology and Centre of Clinical Neuroscience Charles University in Prague, Czech Republic
[10] Neurology Service, Hospital Clínic Universitari, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona; Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Barcelona, Spain
[11] Department of Neurology, Parkinson centre Milano, Italy
[12]Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Rehabilitation, The Netherlands
All on behalf of the European MSA Study Group (EMSA-SG)*

Character count for the title (including spaces and punctuation): 86
Number of references: 30
Number of tables: 2
Number of figures: 1
Total number of words article**:** 3210
Total number of words abstract: 242

Supplemental Data : supplement 1: Physical examination

Corresponding author
Prof. Bastiaan R. Bloem, MD, PhD
Radboud University Medical Centre
PO Box 9101, 6500 HB Nijmegen
The Netherlands
E-mail: bas.bloem@radboudumc.nl

Statistical Analysis conducted by Drs. Borm, MD and R. donders PhD, Radboud University Medical centre

Keywords: Parkinson's disease, atypical parkinsonian disorders, Multiple System Atrophy, parkinsonian type, Progressive Supranuclear Palsy , postural Instability and gait disability.

Abstract

OBJECTIVE: Differentiating Parkinson's disease (PD) from atypical parkinsonian disorders (APD) such as Multiple System Atrophy , parkinsonian type (MSA-p) or Progressive Supranuclear Palsy (PSP-RS) can be challenging. Early signs of postural Instability and gait disability (PIGD) are considered clues that may signal presence of APD. However, it remains unknown which PIGD test – or combination of tests – can best distinguish PD from APD. We evaluated the discriminative value of several widely-used PIGD tests, and aimed to develop a short PIGD evaluation that can discriminate parkinsonian disorders.

METHODS: In this multicentre cohort study patients were recruited by 11 European MSA Study sites. Patients were diagnosed using standardized criteria. Postural instability and gait disability was evaluated using interviews and several clinical tests.

RESULTS: Nineteen PD, 21 MSA-p and 25 PSP-RS patients were recruited. PIGD was more common in APD compared to PD. There was no significant difference in axial symptoms between PSP-RS and MSA-p, except for self-reported falls (more frequent in PSP-RS patients). The test with the greatest discriminative power to distinguish APD from PD was the ability to perform tandem gait (AUC 0.83;95% CI 71-94; p<0.001), followed by the retropulsion test (AUC 0.8; 95% CI 0.69-0.91; p<0.001) and timed-up-and-go test (TUG) (AUC 0.77;95% CI 0.64-0.9; p=0.001). The combination of these three tests yielded highest diagnostic accuracy (AUC 0.96;95% CI 0.92-1.0; p<0.001).

CONCLUSIONS: Our study suggests that simple "bedside" PIGD tests – particularly the combination of tandem gait performance, TUG and retropulsion test – can discriminate APD from PD.

Introduction

Differentiating Parkinson's disease (PD) from atypical parkinsonian disorders (APD) such as the parkinsonian subtype of Multiple System Atrophy (MSA-p) or the classic Richardson's Syndrome presentation of Progressive Supranuclear Palsy (PSP-RS) can be challenging due to overlap in clinical presentation. In early disease stages, patients with PSP-RS and MSA-p are therefore often misdiagnosed.[1, 2] This is detrimental, because an accurate diagnosis is important for patient counselling, for instalment of treatment and to conduct proper research. There are clinical clues (often termed 'red flags') that can help differentiate between PD and APD. Some clues can occur in PSP-RS or MSA-p, but less commonly in PD.[3] The presence of postural instability or gait difficulties (PIGD) in early disease stages is such a clue, because these signs are seen regularly in early stages of APD, but only much later in the course of PD. Several simple tests can detect the presence of PIGD; examples include questions (can you still ride a bicycle?) or several gait and balance tests, such as the retropulsion test, single leg stance test, and timed-up-and-go (TUG) test.[4-8] However, it is unknown which gait and balance test, or which combination of tests, is most sensitive in discriminating between the early stages of PD and atypical parkinsonism (PSP-RS and MSA-p). It is also unknown whether these tests can discriminate between PSP-RS and MSA-p. To address these questions, we performed a multicentre study exploiting the European MSA Study Group network (www.emsa-sg.org), with the primary aim of evaluating the discriminative value of several gait and balance tests. Our secondary aim was to develop a short test battery for gait and balance evaluation that might be used as a simple bedside screening to discriminate PD from APD.

Methods

Participants and setting

Between September 2011 and August 2013, consecutive outpatients were enrolled prospectively at 11 EMSA centres. Limited follow-up was done in small subgroups, due to considerable numbers of dropouts. Here, we solely report on the cross-sectional data. Patients were diagnosed according to the UK Brain Bank[9], revised MSA consensus criteria[10] and NINDS-SPSP criteria.[11] Additional inclusion criteria were: age 30-90 years, ability to walk unassisted, and stable doses of dopaminergic replacement therapy for at least four weeks prior to examination. Exclusion criteria were: secondary cause of parkinsonism as detected by investigation(e.g. vascular parkinsonism, following accepted current criteria[12]*),* co-morbidities that substantially influenced the clinical presentation of PIGD, as judged by the local investigator (e.g. severe polyneuropathy); Hoehn and Yahr stage 4 or 5; prior brain surgery; presence of dementia or major depressive or psychotic disorder. In total 19 PD-patients, 21 MSA-p patients and 25 PSP-RS patients were enrolled. The study was performed in accordance with the declaration of Helsinki. Each study centre obtained local ethical approval. All participants gave their written informed consent prior to participation.

General assessment of motor and cognitive functions

A general assessment of motor and cognitive functions. In PD patients, disease severity was measured using the MDS-UPDRS.[13] The UMSARS was used in MSA-p patients[14] and PSP-Rating Scale was applied in PSP-RS patients.[15] Functional independence in daily living was rated with the Schwab & England scale; a score of 100% indicates a completely independent individual, and 0% indicates complete dependence. Cognitive assessment included the Frontal Assessment Battery (FAB), Mini Mental State Examination (MMSE) and Beck Depression Inventory (BDI).[16]

Gait and balance assessment

PIGD signs were evaluated using both history taking and physical examination. History taking included several questions that could be answered with yes or no. If patients were unable to answer the question, they were excluded for the corresponding analysis. We evaluated the following items:[8, 13, 17-21]

- History of Freezing of Gait (FOG): participants were asked whether they had ever experienced FOG. FOG was explained as 'feeling as if the feet were being glued to the floor, hampering effective forward stepping'.
- Postural instability: we asked participants whether they were easily lost balance (with or without falling), e.g. during gait initiation or turning.

- Falling: we asked participants if they had fallen over the past 6 months. A fall was considered an event during which the individual had unintentionally come to rest on the ground or other lower level.
- Difficulties arising from a chair: we asked participants whether they experienced difficulties arising from a chair, e.g. needing multiple attempts, being slow and clumsy, or unable to arise without using their arms.
  Bicycle sign: participants were asked if they could still ride a bicycle, or whether they had stopped riding a bicycle because of self-perceived balance impairment.

Physical examination consisted of the following tests (see supplement 1 for details), performed in the following order:

- Walking velocity, both under a single and dual tasking condition.
- Timed-Up-and-Go-test (TUG): the time needed to stand up from a chair with arms, walk 3 meters at a comfortable pace, turn around, walk back to the chair and sit down.
- FOG: rapid 360° turns, and walking fast with short steps.
- Tandem gait (10 consecutive steps along a straight line at self-preferred speed, without support and with eyes open). This trial was performed with and without an actual visually line
- Single Leg Stance test (SLS): standing on one leg with hands held onto the hips, and with contralateral foot floating in the air, for as a long as possible or for a maximum duration of 30 seconds.
- Retropulsion test.

All gait and balance items were videotaped and rated by a non-blinded experienced rater (CB). Participants were tested regardless of their medication dose cycle. Walking aids were not allowed.

Statistical analysis

ANOVA with alpha level 0.05 was used to determine which variables (age, disease duration, gender, Schwab & England score, MMSE, FAB, BDI, walking velocity, dual tasking, SLS, TUG and retropulsion test) and diagnosis (PD, MSA-p and PSP-RS) differed across groups. ANOVA was followed by post-hoc multiple comparison procedure Dunnett's T3 test for the PIGD items, to identify differences between diagnosis groups whilst correcting for unequal variances. Pearson's chi-square test or Fisher's exact test were used to analyse nominal variables (levodopa benefit, tandem gait with/without cue, SLS <10 seconds, TUG >16 seconds, history of FOG, falls in last 6 months, postural instability, difficulties arising from chair, bicycle sign, stopped walking when talking, and FOG), choosing Fisher's exact test when criteria for Pearson's chi-square were not obtained. To evaluate the diagnostic effect size of PIGD tests, nonparametric receiver operating characteristics (ROC) curves were plotted and areas under curve (AUC) were calculated. Cut-off points were estimated using ROC curves for the TUG and SLS test.

We first summarized PSP-RS and MSA-p patients under the umbrella term APD, and compared this group with PD, since this distinction is clinically most relevant. We then compared the MSA-p and PSP-RS groups. Finally, we included significant predictors found using the ROC's in a multivariable binary logistic regression model, to test the discriminative value of these predictors together in differentiating APD from PD. Subjects with missing items were excluded from the corresponding analysis.

Results

Characteristics

PD, PSP-RS and MSA-p groups did not differ with respect to age, gender and disease duration (Table 1). Disease duration was calculated using year of symptom onset. PD subjects showed greater independence in daily living than patients with APD, as reflected by the Schwab & England scores (PD 88%; MSA-p 68%; PSP-RS 66%; $F_{2,58}$=17.5, p<.001). The proportion of subjects reporting levodopa benefit was significantly higher among PD patients. (. MMSE-scores were comparable across groups. PSP patients had more pronounced frontal executive dysfunction, as indicated by lower FAB scores (mean 11.96; standard deviation [SD] 4.08), compared to MSA-p patients (15.81; SD 2.64) and PD patients (16.16; SD 2.19) ($F_{2,61}$=12.2, p<.001). The

Depression Inventory was lower in PD patients (mean 8.44; SD 5.82) compared to patients with MSA-p (12.86; SD 5.63) and PSP-RS (14.00; SD 8.67) ($F_{2,58}$=3.5, p=.04). Post-hoc Dunnett t3 tests showed a significant difference between PSP-RS and MSA-p (p=.001).

Postural instability and gait disability profiles

Table 2 shows the results for PIGD items. Twenty-two percent of PD subjects reported postural instability, compared to 91% of PSP-RS patients, and 80% of MSA-p patients (χ^2(2, N=60)= 23.42, p<.001). Difficulties arising from a chair were significantly less common in PD patients than in APD: 17% of PD patients reported difficulties in contrast to 77% of PSP-RS patients and 65% of MSA-p patients (χ^2(2, N=60)= 1590, p<.001). Half of PD patients were still able to ride a bicycle, in contrast to 17% of PSP-RS patients (χ^2 (1, N =32) = 4.07, p=0.044), and 6% of MSA-p patients (χ^2(1, N=30)=7.31, p=.012). Almost all PSP-RS patients reported falls in the past 6 months (96%), whereas only 17% of the PD group reported a fall. Falls were also common in MSA-p patients, with 60% reporting falls (χ^2(2, N =61)= 26.40, p<.001). PD subjects showed a faster mean walking velocity compared to PSP-RS patients and MSA-p patients, both with a dual task ($F_{2,58}$=6.5, p=.003) and without ($F_{2,55}$=7.1, p=.002). There was no group difference for the frequency of "stops with walking" during dual tasking. Patients with PD could stand longer on one leg during the SLS-test (17 sec; SD 10.8) compared to MSA-p patients (7.4 sec; SD 6.9) and PSP-RS patients (7.2 sec; SD 7.7) ($F_{2,55}$=8.4, p=.001). Among PD patients, 60% could stand on one leg for 10 seconds, compared to only 25% of PSP-RS patients and 25% of MSA-p patients. Retropulsion test was normal in 83% of PD patients, compared to only 28% of PSP-RS patients and 44% of MSA-p patients (χ^2(2, N=63)=15.45, p<.001). The same pattern was seen for tandem gait performance (no line). Tandem gait was undisturbed (no single side-step) in 88% of PD patients, compared to only 15% of MSA-p patients and 30% of PSP-RS patients (χ^2(2, N=60)=22.21, p< 0.001). Comparable results were observed when tandem gait was performed with a line visually present (McNemar test, p=.344 (2 sided)). The time to complete the TUG-test was shorter in PD patients (12 sec; SD 3.4) compared to PSP-RS patients (22 sec; SD 15.4) and MSA-p patients (18 sec; SD 6.5) ($F_{2,57}$=5.4, p=.001). Most PD patients (98%) completed the TUG-test within 16 seconds, compared to only 26% of MSA-p patients and 40% of PSP-RS patients (χ^2(2, N=61)= 14.90, p=.001). (Table 2)

Diagnostic accuracy

Figure 1 presents the AUC and 95% CI for the PIGD items when comparing PD vs APD. The history taking items that best discriminated between PD and APD were presence of subjective postural instability (AUC 0.82; 95% CI 69-94; p<.001) and falls in the past 6 months (AUC 0.81; 95% CI 69-95; p<.001). The physical exam tests that best discriminated PD and APD were tandem gait performance (no line; AUC 0.83; 95% CI 71-94; p<.001), followed by the TUG-test (AUC 0.83; 95% CI 72-94; p<.001). For clinical use, we estimated cut-off points for the TUG-test and the SLS-test, based on the optimal AUC curve. The most discriminative cut-off for the TUG-test was 16 seconds (AUC 0.77; 95% CI 64-90; p=.001) and 20 seconds for the SLS-test (AUC 0.71; 95% CI 55-86; p=.012). The retropulsion test yielded an AUC of 0.80 (95% CI 69-91). Logistic regression showed that discrimination rates were highest for the combination of TUG (cut-off 16 sec), tandem gait test (no line) and retropulsion test, with an AUC of 0.96 (95% CI 0.92-1.0; p<.001). Exchanging the retropulsion test for a history of subjective postural instability yielded an AUC of 0.95 (95% CI 0.92-1.0; p<.001).

Discussion

This study evaluated primarily which gait and balance tests might help to differentiate PD from two forms of APD (PSP-RS and MSA-p). We also investigated which tests might differentiate between MSA-p and PSP-RS. Our results show that PIGD is more common in patients with MSA-p and PSP-RS than in patients with PD. The most sensitive single test to discriminate between PD and APD was the tandem gait test. Other discriminative tests included the TUG-test and retropulsion test. The combination of tandem gait test, TUG-test (cut-off 16 sec) and retropulsion test yielded the highest diagnostic accuracy (AUC 0.96). Gait and balance tests could not discriminate between MSA-p and PSP-RS.

These results cannot be interpreted without considering possible group differences. The three groups did not differ with respect to gender, age and disease duration. Not surprisingly, the APD groups had greater disability, and PSP-RS patients had more pronounced frontal executive dysfunction compared to both MSA-p and PD patients. Also, scores for mood were better for PD patients compared to both MSA-P patients and, particularly, PSP-RS patients. Importantly, all patients already fulfilled established clinical diagnostic criteria for PD, MSA-p and PSP-RS, so the test battery identified here essentially only confirmed what clinicians had already decided based upon the overall clinical presentation (which obviously also included some PIGD items). The present findings therefore need to be replicated in a cohort of patients in earlier disease stages (mean disease duration for the present cohort was 4 years, calculated since onset of first symptoms), when there is greater diagnostic uncertainty, to see if early detection of specific PIGD abnormalities can accelerate the differential diagnostic process. It would also be interesting to see if our test battery can reflect the degree of postural instability (it currently only evaluates the presence or absence of postural instability but does not yet accurately rate the severity).
Clinicians currently use a variety of tests to evaluate the presence of PIGD-abnormalities, and this likely takes place very inconsistently across clinics, dictated by local customs, personal preferences, prior experience and perhaps time constraints. Here, we show that the combined performance of the tandem gait test, TUG-test and retropulsion test improves the sensitivity and at the same time avoids diagnostically superfluous, time-consuming, expensive and potentially harmful investigations. Obviously, the findings of our proposed test battery should never be judged in isolation during the diagnostic process. Implementation of the entire test battery as one part of an otherwise full exam might possibly lead to an accurate diagnosis in an earlier disease stages, which is of importance for patient counselling (patients are keen to know which disease they have, even when this does not have immediate treatment consequences) and inclusion in clinical trials (namely better exclusion of patients with atypical parkinsonism in trials that aim to include patients with PD, and vice versa).

Tandem gait performance was the most sensitive test to separate PD from APD, with a sensitivity of 77% and specificity of 88% (AUC 0.83). This high discriminative value is in line with previous studies reporting a sensitivity ranging from 82-90.6%, a specificity ranging from 66-92% and AUC of 0.81.[7, 22, 23] The low specificity of 66.6% in one study[23] might have been related to the longer disease duration of participants in that study compared to the present one (8.4 vs 4 years). The new finding here is that the tandem gait test outperformed a wide range of other established PIGD questions and tests. Impaired tandem gait performance signals the presence of medio-lateral balance impairment, which is typically absent in early PD, but might develop later as the neurodegenerative process spreads.[18] In contrast, this more widespread neurodegeneration is already present in early stages of PSP-RS and MSA-p.[24, 25] Another new finding was the refinement of how tandem gait should be executed in clinical practice, to yield the best diagnostic value. Our results suggest that tandem gait performance can reveal medio-lateral balance impairment, independent of the presence of a visual reference on the floor. The TUG-test was the second-best physical examination test to discriminate between PD and APD. Previous studies used the TUG-test mainly for a different purpose, namely to predict fall risks, and mainly in elderly populations.[26] In the PD patients tested here, a long TUG time ($\geq$16 sec) was also associated with an increased fall risk (OR 3.86, 95% CI 1.05,14.27, P=0.043), and this is in line with earlier work.[20] Various studies have emphasized that patients with PD might be at risk of developing concurrent cerebrovascular lesions, and that such lesions may be clinically relevant.[27-29]Other patients may have vascular parkinsonism, i.e. parkinsonian signs that can be fully attributed to underlying cerebrovascular pathology.[12] It would be interesting to see if the TUG, retropulsion test and tandem gait performance also distinguish between PD and vascular parkinsonism, or whether development of concurrent cerebrovascular lesions in patients with otherwise typical PD might influence performance on the PIGD test battery.

Our present findings also confirm the discriminative value of the 'bicycle sign' (having stopped riding a bicycle because of self-perceived instability), with an AUC (0.69 in the present study) that was largely comparable to a previous study (which reported an AUC of 0.74).[17] This is interesting, because the earlier study was performed in the Netherlands, where cycling is very common, even at older ages. The present

study was performed not only in the Netherlands, but also in various different European countries, including those where cycling is not an integral part of the everyday lifestyle of patients. This suggests that the bicycle sign is useful test even outside the Netherlands, as was also suggested by another study performed in Japan.[30] Note that the bicycle sign actually tests the same construct of medio-lateral instability as the tandem gait test. Tandem walking can serve as a simple alternative for patients who have never cycled in their life.

The combination of tandem gait test, TUG-test (cut-off 16s) and retropulsion test yielded the highest discriminative value. This specific combination is understandable because the three test are complementary, each exploring different constructs: tandem gait performance assesses medio-lateral instability; the TUG-test assesses transfers in and out of a chair with arms, walking and turning; and the retropulsion test assesses anterior-posterior instability. We suggest to use this combination as a short and simple bedside test battery to discriminate PD from APD. It is possible to exchange the retropulsion test for a history of subjective postural instability, as this combination also yielded a good accuracy (AUC 0.95). One advantage is that asking about subjective postural instability is less cumbersome than performing a retropulsion test. A disadvantage to asking about prior falls is over- or underreporting by the patient. We know that asking about prior falling is rather susceptible to recall bias. [31] History taking of postural instability and falling are not complementary, as both explore the same construct of anterior-posterior instability. Note that several commonly used tests did not contribute much to the differential diagnosis. This included the 360 degrees rapid turning test (to elicit FOG)[32, 33]), which was equally impaired in PD, MSA-p and PSP-RS. Standing on one leg was more frequently impaired among MSA-p and PSP-RS patients compared to PD patients, but added no further discriminatory value to the battery described above. Finally, the “stops walking while talking” – which is a good test to predict the risk of future falls – did not differentiate between the three groups.

Our results show that presence or absence of PIGD cannot differentiate between MSA-p and PSP-RS. Both patient groups were comparable for all investigated items, except for presence of falls in the last 6 months, which was higher in PSP-RS patients compared to MSA-P patients. However, the mere presence of prior falls seems insufficient to reliably differentiate between MSA-p and PSP-RS.[31] For this purpose, careful evaluation of additional signs are needed, such as presence of supranuclear vertical gaze palsy or prominent early cognitive impairment (supporting a diagnosis of PSP-RS), and marked early autonomic dysfunction (supporting a diagnosis of MSA-p).[11, 34]

Several limitations must be mentioned. Our study size was relatively small, hence the regression analysis might have generated an overestimate due to limited statistical power. Additionally, the video-rater was not blinded, and future studies should preferably apply blinded video-rating. Moreover a rate of misdiagnosis should be considered since we have no post-mortem examinations. As an acceptable surrogate for a definite neuropathological diagnosis, we used the clinical diagnosis by experts in movement disorders. Future studies remain needed to confirm the discriminative and longitudinal value of our bedside test battery, in particular in patients in very early disease stages who do not yet fulfil established criteria for a particular form of parkinsonism, and where the diagnostic uncertainty is greatest. Furthermore, all patients in the present study were tested while ON medication, the advantage being that this is how the large majority of patients will be tested in busy regular clinical practices. However, given the fact that the responsiveness to dopaminergic medication varied across the groups, this choice for testing in an ON state could have affected the results. We suspect that the influence of medication may not have been large, because many axial parkinsonian features tend to be largely resistant to dopaminergic medication. To examine this effect of medication on the discriminatory power, it would be interesting to test patients both OFF and ON medication.

Our study focused primarily on separating the group of atypical parkinsonisms from PD, realizing that distinguishing between the various forms of atypical parkinsonism is much more difficult. This certainly applies to the often-great difficulties in distinguishing PSP-RS from PSP-P, particularly in early disease

stages. We acknowledge that our study is of limited value to make a differentiation between subgroups of parkinsonism, and we encourage that this issue will be addressed in future studies. Finally, we acknowledge that there may be a difference in PIGD profile between PD subgroups (e.g. tremor predominant versus non-tremor predominant) and PSP subgroups (e.g. PSP-RS versus PSP-P), which we encourage to be addressed in future studies.

Author contributions

Drs. Carlijn Borm - Study concept and design, acquisition of data, - analysis and interpretation, drafting the manuscript.
Dr. Florian Krismer - Study concept and design - acquisition of data- critical revision of the manuscript for important intellectual content
Dr. Gregor Wenning - Study concept and design - study supervision- critical revision of the manuscript for important intellectual content
Dr. Klaus Seppi - critical revision of the manuscript for important intellectual content
Dr. Werner Poewe - critical revision of the manuscript for important intellectual content
Dr. Maria Teresa Pellecchia- acquisition of data
Dr. Paolo Barone- critical revision of the manuscript for important intellectual content
Dr. Erik Lisbjerg Johnsen - acquisition of data
Dr. Karen Østergaard- critical revision of the manuscript for important intellectual content
Dr. Tanya Gurevich - acquisition of data, critical revision of the manuscript for important intellectual content
Dr. Ruth Djaldetti - acquisition of data, critical revision of the manuscript for important intellectual content
Dr. Luisa Sambati - acquisition of data
Dr. Pietro Cortelli- critical revision of the manuscript for important intellectual content
Dr. Igor Petrović - acquisition of data
Dr. Vladimir s. Kostić - critical revision of the manuscript for important intellectual content
Dr. Hana Brožová - acquisition of data
Dr. Evžen Růžička - critical revision of the manuscript for important intellectual content
Dr. Maria Jose Marti - acquisition of data
Dr. Eduardo Tolosa - critical revision of the manuscript for important intellectual content
Dr. Margherita Canesi - acquisition of data
Dr. Bart Post - acquisition of data- critical revision of the manuscript for important intellectual content
Dr. Jorik Nonnekes – study supervision- critical revision of the manuscript for important intellectual content
Dr. Bastiaan R. Bloem - Study concept and design- study supervision - critical revision of the manuscript for important intellectual content

*Members of the European MSA Study Group:

Aarhus, Denmark, Dr. Karen Østergaard; Athens, Greece, Maria Stamelou Md,PhD; Barcelona, Spain, Prof. Eduardo Tolosa, Belgrade, Serbia, Prof. Vladimir S. Kostic; Bologna, Italy, Prof. Pietro Cortelli, MD; Bonn, Germany: Prof. Thomas Klockgether, Dr. Richard Dodel, Dr. Michael Abele; Bordeaux, France: Prof. Wassilios Meissner; Dresden, Germany, Heinz Reichmann, MD,PhD; Dublin, Ireland, prof. Tim Lynch; Gdansk, Poland prof. Jaroslaw Slawek; Innsbruck, Austria: Prof. Werner Poewe, Prof. Gregor K. Wenning, Mag. Dr. Klaus Seppi, Dr. Florian Krismer; Kiel, Germany: Prof. Daniela Berg, Lisbon, Portugal: Prof. Joaquim Ferreira; London, United Kingdom: Prof. Henry Houlden, Niall P. Quinn; Lund, Sweden:, Prof. Håkan Widner, Prof; Manchester, United Kingdom dr. Alexander Gerhard, MD; Marburg, Germany:, Dr. Karla Maria Eggert; Milan, Italy: Prof. Alberto Albanese, Dr. Francesca del Sorbo; Naples, Italy: Prof. Paolo Barone, Dr. Maria T. Pellecchia; Nijmegen, the Netherlands; Prof. Bas Bloem, Carlijn Borm MD; Petach- Tiqva, Israel: Dr. Ruth Djaldetti; Rome, Italy: Prof. Alfredo Berardelli, Prof. Carlo Colosimo; Santander, Spain: Prof. Jose Berciano; Sofia, Bulgaria, prof Latchezar Traykov; Tel-Aviv, Israel: Dr. Nir Giladi, Dr. Tanya Gurevich; Toulouse, France: Prof. Olivier Rascol, Dr. Monique Galitzky, Tübingen Germany: Prof. Thomas Gasser,

Acknowledgements
R. Donders supervision of statistical analysis.
The Stichting Parkinson Fonds for funding.

Author Disclosures:
Carlijn D.J.M. Borm: Reports no disclosures.

Florian Krismer: Reports no disclosures
Gregor K. Wenning : Reports no disclosures.
Klaus Seppi: Reports no disclosures
Werner Poewe: Reports no disclosures.
Maria Teresa Pellecchia : Reports no disclosures.
Paolo Barone : Reports no disclosures.
Erik Dupont: Reports no disclosures.
Karen Østergaard : Reports no disclosures.
Tanya Gurevich: Reports no disclosures.
Margherita Canesi : Reports no disclosures.
Ruth Djaldetti: Reports no disclosures.
Luisa Sambati: Reports no disclosures.
Pietro cortelli : Reports no disclosures.
Igor Petrović : Reports no disclosures.
Vladimir Kostić: Reports no disclosures.
Hana Brožová: Reports no disclosures.
Evžen Růžička : Reports no disclosures.
Eduardo Tolosa : Reports no disclosures.
Bart Post : Reports no disclosures.
Jorik Nonnekes: Reports no disclosures.
Bas Bloem: Reports no disclosures.

Study funded by: The Stichting Parkinson Fonds *(grant number 38000)*

None of the funding sources had any influences on the study protocol, interpretation of results, writing of this paper or decision to submit the paper for publication.

References

[1] A.J. Hughes, S.E. Daniel, A.J. Lees, Idiopathic Parkinson's disease combined with multiple system atrophy. A clinicopathological report, Movement disorders : official journal of the Movement Disorder Society 6(4) (1991) 342-6.
[2] R.B. Postuma, D. Berg, M. Stern, W. Poewe, C.W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A.E. Lang, G. Halliday, C.G. Goetz, T. Gasser, B. Dubois, P. Chan, B.R. Bloem, C.H. Adler, G. Deuschl, MDS clinical diagnostic criteria for Parkinson's disease, Movement disorders : official journal of the Movement Disorder Society 30(12) (2015) 1591-601.
[3] M. Kollensperger, F. Geser, K. Seppi, M. Stampfer-Kountchev, M. Sawires, C. Scherfler, S. Boesch, J. Mueller, V. Koukouni, N. Quinn, M.T. Pellecchia, P. Barone, N. Schimke, R. Dodel, W. Oertel, E. Dupont, K. Ostergaard, C. Daniels, G. Deuschl, T. Gurevich, N. Giladi, M. Coelho, C. Sampaio, C. Nilsson, H. Widner, F.D. Sorbo, A. Albanese, A. Cardozo, E. Tolosa, M. Abele, T. Klockgether, C. Kamm, T. Gasser, R. Djaldetti, C. Colosimo, G. Meco, A. Schrag, W. Poewe, G.K. Wenning, M.S.A.S.G. European, Red flags for multiple system atrophy, Movement disorders : official journal of the Movement Disorder Society 23(8) (2008) 1093-9.
[4] S. Morris, M.E. Morris, R. Iansek, Reliability of Measurements Obtained With the Timed “Up & Go” Test in People With Parkinson Disease, Physical therapy 81 (2001) 810-818.
[5] J.V. Jacobs, F.B. Horak, V.K. Tran, J.G. Nutt, Multiple balance tests improve the assessment of postural stability in subjects with Parkinson's disease, Journal of neurology, neurosurgery, and psychiatry 77(3) (2006) 322-6.
[6] J. Nonnekes, R. Goselink, V. Weerdesteyn, B.R. Bloem, The retropulsion test: a good evaluation of postural instability in Parkinson's disease?, Journal of Parkinson's disease 5(1) (2015) 43-7.
[7] W.F. Abdo, G.F. Borm, M. Munneke, M.M. Verbeek, R.A. Esselink, B.R. Bloem, Ten steps to identify atypical parkinsonism, Journal of neurology, neurosurgery, and psychiatry 77(12) (2006) 1367-9.
[8] M.K. Mak, M.Y. Pang, Balance confidence and functional mobility are independently associated with falls in people with Parkinson's disease, Journal of neurology 256(5) (2009) 742-9.
[9] D.J. Gelb, E. Oliver, S. Gilman, Diagnostic criteria for Parkinson disease, Archives of neurology 56(1) (1999) 33-9.
[10] S. Gilman, P.A. Low, N. Quinn, A. Albanese, Y. Ben-Shlomo, C.J. Fowler, H. Kaufmann, T. Klockgether, A.E. Lang, P.L. Lantos, I. Litvan, C.J. Mathias, E. Oliver, D. Robertson, I. Schatz, G.K. Wenning, Consensus statement on the diagnosis of multiple system atrophy, J Auton Nerv Syst 74(2-3) (1998) 189-92.
[11] I. Litvan, Y. Agid, D. Calne, G. Campbell, B. Dubois, R.C. Duvoisin, C.G. Goetz, L.I. Golbe, J. Grafman, J.H. Growdon, M. Hallett, J. Jankovic, N.P. Quinn, E. Tolosa, D.S. Zee, Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop, Neurology 47(1) (1996) 1-9.
[12] J.C. Zijlmans, S.E. Daniel, A.J. Hughes, T. Revesz, A.J. Lees, Clinicopathological investigation of vascular parkinsonism, including clinical criteria for diagnosis, Movement disorders : official journal of the Movement Disorder Society 19(6) (2004) 630-40.
[13] C.G. Goetz, B.C. Tilley, S.R. Shaftman, G.T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M.B. Stern, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A.E. Lang, A. Lees, S. Leurgans, P.A. LeWitt, D. Nyenhuis, C.W. Olanow, O. Rascol, A. Schrag, J.A. Teresi, J.J. van Hilten, N. LaPelle, U.R.T.F. Movement Disorder Society, Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results, Movement disorders : official journal of the Movement Disorder Society 23(15) (2008) 2129-70.
[14] G.K. Wenning, F. Tison, K. Seppi, C. Sampaio, A. Diem, F. Yekhlef, I. Ghorayeb, F. Ory, M. Galitzky, T. Scaravilli, M. Bozi, C. Colosimo, S. Gilman, C.W. Shults, N.P. Quinn, O. Rascol, W. Poewe, G. Multiple System Atrophy Study, Development and validation of the Unified Multiple System Atrophy Rating Scale (UMSARS), Movement disorders : official journal of the Movement Disorder Society 19(12) (2004) 1391-402.
[15] L.I. Golbe, P.A. Ohman-Strickland, A clinical rating scale for progressive supranuclear palsy, Brain : a journal of neurology 130(Pt 6) (2007) 1552-65.
[16] M.F. Folstein, S.E. Folstein, P.R. McHugh, "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician, J Psychiatric Research 12(3) (1975) 189-198.

[17] M.B. Aerts, W.F. Abdo, B.R. Bloem, The "bicycle sign"for atypical parkinsonism, The lancet 377 (2011) 125-126.
[18] J. Nonnekes, M.B. Aerts, W.F. Abdo, B.R. Bloem, Medio-lateral balance impairment differentiates between Parkinson's disease and atypical parkinsonism, Journal of Parkinson's disease 4(4) (2014) 567-9.
[19] B.R. Bloem, Y.A. Grimbergen, M. Cramer, R.A. Roos, Prospective assessment of falls in Parkinson's disease Movement disorders : official journal of the Movement Disorder Society 10 (1998) 147.
[20] M.K. Mak, M.Y. Pang, Parkinsonian single fallers versus recurrent fallers: different fall characteristics and clinical features, Journal of neurology 257(9) (2010) 1543-51.
[21] B.R. Bloem, V.V. Valkenburg, M. Slabbekoorn, J.G. Dijk van, The multiple tasks test. strategies in Parkinson's disease, exp brain res 137 (2001) 478-486.
[22] M.B. Aerts, R.A. Esselink, W.F. Abdo, F.J. Meijer, G. Drost, N. Norgren, M.J. Janssen, G.F. Borm, B.R. Bloem, M.M. Verbeek, Ancillary investigations to diagnose parkinsonism: a prospective clinical study, Journal of neurology 262(2) (2015) 346-56.
[23] H. Morales-Briceno, M. Rodriguez-Violante, D. Martinez-Ramirez, A. Cervantes-Arriaga, A reappraisal of the ten steps test for identifying atypical parkinsonism, Clin Neurol Neurosurg 119 (2014) 1-3.
[24] B.R. Bloem, K.P. Bhatia, Gait and balance in basal ganglia disorders, Clinical Disorders of Balance, Posture and Gait. (173 - 206). (2004).
[25] D.C. Paviour, S.L. Price, M. Jahanshahi, A.J. Lees, N.C. Fox, Regional brain volumes distinguish PSP, MSA-P, and PD: MRI-based clinico-radiological correlations, Movement disorders : official journal of the Movement Disorder Society 21(7) (2006) 989-96.
[26] O. Beauchet, B. Fantino, G. Allali, S.W. Muir, M. Montero-Odasso, C. Annweiler, Timed Up and Go test and risk of falls in older adults: a systematic review, The journal of nutrition, health & aging 15(10) (2011) 933-8.
[27] B.L. Ten Harmsen, A. van Rumund, M.B. Aerts, M.I. Bergkamp, R.A.J. Esselink, E. Richard, F.J.A. Meijer, B.R. Bloem, D.J. van Wamelen, Clinical correlates of cerebral white matter abnormalities in patients with Parkinson's disease, Parkinsonism Relat Disord 49 (2018) 28-33.
[28] I. Rektor, D. Goldemund, K. Sheardova, I. Rektorova, Z. Michalkova, M. Dufek, Vascular pathology in patients with idiopathic Parkinson's disease, Parkinsonism Relat Disord 15(1) (2009) 24-9.
[29] N.I. Bohnen, M.L. Muller, N. Zarzhevsky, R.A. Koeppe, C.W. Bogan, M.R. Kilbourn, K.A. Frey, R.L. Albin, Leucoaraiosis, nigrostriatal denervation and motor symptoms in Parkinson's disease, Brain : a journal of neurology 134(Pt 8) (2011) 2358-65.
[30] H. Miwa, T. Kondo, Bicycle sign for differential diagnosis of parkinsonism: is it of use in a hilly country like Japan?, Journal of Parkinson's disease 1(2) (2011) 167-8.
[31] S.R. Cummings, M.C. Nevitt, S. Kidd, Forgetting falls. The limited accuracy of recall of falls in the elderly, Journal of the American Geriatrics Society 36(7) (1988) 613-6.
[32] A.H. Snijders, M.J. Nijkrake, M. Bakker, M. Munneke, C. Wind, B.R. Bloem, Clinimetrics of freezing of gait, Movement disorders : official journal of the Movement Disorder Society 23 Suppl 2 (2008) S468-74.
[33] J. Spildooren, S. Vercruysse, K. Desloovere, W. Vandenberghe, E. Kerckhofs, A. Nieuwboer, Freezing of gait in Parkinson's disease: the impact of dual-tasking and turning, Movement disorders : official journal of the Movement Disorder Society 25(15) (2010) 2563-70.
[34] A. Lipp, P. Sandroni, J.E. Ahlskog, R.D. Fealey, K. Kimpinski, V. Iodice, T.L. Gehrking, S.D. Weigand, D.M. Sletten, J.A. Gehrking, K.K. Nickander, W. Singer, D.M. Maraganore, S. Gilman, G.K. Wenning, C.W. Shults, P.A. Low, Prospective differentiation of multiple system atrophy from Parkinson disease, with and without autonomic failure, Archives of neurology 66(6) (2009) 742-50.

Table 1 Characteristics

	PD (*n*=19)	**PSP-RS** (*n*=25)	**MSA-p** (*n*=21)	*** P-value**
	mean (range)	mean (range)	mean (range)	
Gender	15m	18m	11m	n.s.
Age at examination (years)	62 (41-81)	67 (57-79)	65 (50-80)	n.s.
Age at symptom onset (years)	56 (33-72)	63 (50-75)	61 (45-72)	n.s.
Disease duration (years)	4 (2-8)	4 (1-8)	4 (1-9)	n.s.
Levodopa benefit	100%(n=12[x])	33% (n=18)	63% (n=16)	0.004[a]
Schwab & England ADL scale	88% (80-100)	66% (40-90)	68% (50-100)	<0.001[a,b]
Rating scales	**PD**	**PSP-RS**	**MSA-p**	
UMSARS I			17 (8-35)	
UMSARS II			19 (12-37)	
MDS-UPDRS I	7 (2-16)			
MDS-UPDRS II	9 (0-17)			
MDS-UPDRS III	24 (9-34)			
PSPRS		36 (15-60)		
Cognitive function assessment	**PD**	**PSP-RS**	**MSA-p**	
MMSE	28 (21-30)	27 (23-30)	28 (14-30)	n.s.
FAB	16 (9-18)	12 (3-18)	16 (9-18)	<0.001[a,c]
BDI	8 (0-23)	14 (0-37)	13 (2-22)	0.039

M: men, Age at examination and age at symptom onset are in years. Activities of daily living scale according to the modified Schwab and England scale (score 0-100%), *UMSARS* Unified Multiple system Atrophy Rating Scale part I (score 0-48), UMSARS part II (score 0-56), *UPDRS MDS* unified Parkinson's disease rating scale part I (0-52), UPDRS part II (score 0-52), UPDRS part III (score 0-132), *PSPRS* Progressive Supranuclear Palsy Rating Scale (score 0-100), *FAB* Frontal Assessment Battery (score 0-18), *MMSE* Mini Mental State Examination (0-30), *BDI* Becks depression inventory (0-63) score (0-13), light (14-19), moderate (20-28), (severe 29-63) M=male

For MDS-UPDRS, UMSARS and PSPRS higher scores indicate worse functioning. For activities of daily living scale, FAB and MMSE lower scores indicate worse functioning. N.S. not statistically significant.

[x] 7 PD patients had never used levodopa, they were therefore scored as unknown and were excluded from this table.

*P-value for the comparison of PD, MSA-P and PSP-RS by analysis of variance (ANOVA) for continuous variables or Pearson's chi square tests for categorical variables. Fisher's exact tests were abducted when rules applying to chi-square tests were not obtained.

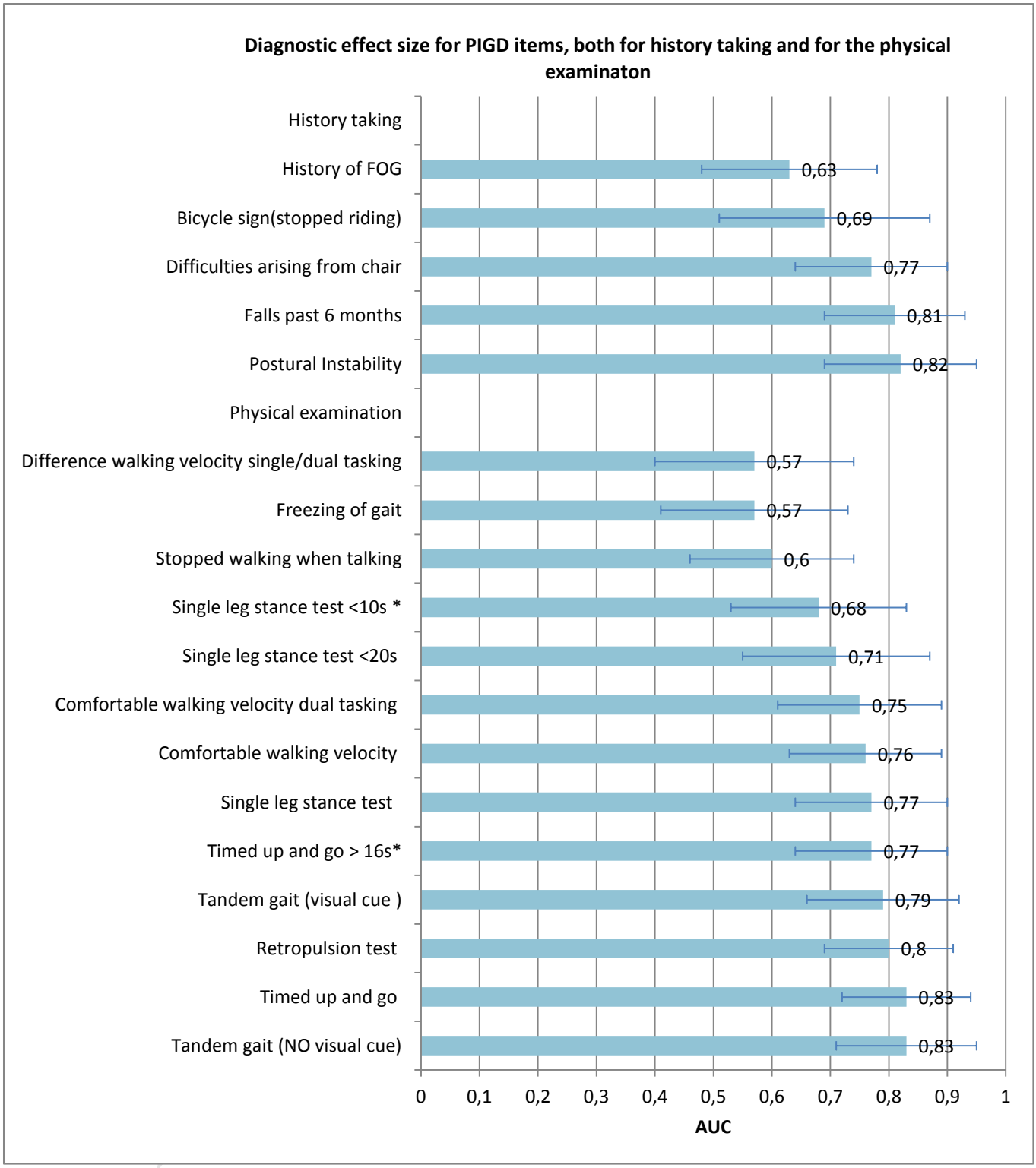
[a] indicates significant difference between PSP and PD p<0.05, [b] indicates significant difference between MSA and PD p<0.05, [c] indicates significant difference between MSA and PSP p<0.05

Table 2 Postural instability and gait disorder items

	PD		PSP-RS		MSA-p		
History	mean ± SD	n=	mean ± SD	n=	mean ± SD	n=	p-value
History of FOG	17%	18	39%	23	45%	20	0.155
Postural Instability	22%	18	91%	22	80%	20	<0.001
Falls in last 6 months	17%	18	96%	23	60%	20	<0.001
Difficulties arising from chair	17%	18	77%	22	65%	20	<0.001
Bicycle sign (stopped riding)	50%	14	83%	18	94%	16	0.019
Physical examination							
Comfortable walking velocity (m/s)	0.8 ± 0.2	18	0.5± 0.3	24	0.6± 0.2	20	0.002
Comfortable walking velocity dual tasking (m/s)	0.6 ± 0.2	18	0.4 ± 0.2	23	0.5± 0.2	20	0.003
Difference walking velocity single and dual tasking	0.17 ± 0.2	18	0.15 ± 0.1	23	0.14 ± 0.1	20	0.870
Difference between single and dual tasking steps	0.17 ± 0.2	18	1.8 ± 4.1	20	2.1 ± 4.2	17	
Stopped walking when talking	16%	19	40%	25	35%	20	0.208
Single leg stance test (s)	16.6 ± 10.8	18	7.4± 6.9	24	7.2±7.7	20	0.001
Single leg stance test < 10s *	39%	18	75%	24	75%	20	0.026
Single leg stance test <20 s	50%	18	96%	24	85%	20	0.001
retropulsion test (score according MDS-UPDRS)	0.2 ± 0.5	18	1.8 ±1.5	25	1.7±1.5	20	0.001
0: normal no problems, 1-2 steps	83%		28%		44%		
1: slight 3-5 steps	11%		26%		22%		
2: mild >5 steps, unaided recovery	6%		0%		2%		
3: moderate: falls if not caught	0%		28%		22%		
4: severe: very unstable	0%		16%		10%		
Timed up and go (s)	12 ±3.4		22.4± 15.4	24	18.0± 6.5	20	0.007
Timed up and go > 16s*	12%	17	71%	24	60%	20	< 0.001
Freezing of Gait	22%	18	32%	25	40%	20	0.501
Tandem gait (NO line present): without side steps	88%	17	30%	23	15%	20	<0.001
Tandem gait (line visually present): without side steps	77%	17	22%	23	15%	20	<0.001

*cut off points as described by[5, 8]. *FOG* freezing of gait

P-value for the comparison of PD, MSA-p and PSP-rs by analysis of variance (ANOVA) for continuous variables or Pearson's chi square tests for the categorical variable. After significant group interactions, post hoc Dunnett's t3 and paired chi square tests were performed. All items were significant (p<0.05) for PD vs MSA or PSP, only falls indicates significant difference between MSA and PSP. Fisher's exact tests were abducted when the rules applying to chi-square tests were not obtained.



Diagnostic effect size between PD and atypical parkinsonism for each PIGD test obtained by area under the curve (AUC). Error bars indicate 95% confidence intervals (CI).

Highlights

- Postural instability is more common in atypical parkinsonism than Parkinson's disease
- Tandem gait performance discriminates between atypical parkinsonism and PD
- Impaired tandem gait performance indicates medio-lateral balance impairment
- Combined tests can improve sensitivity and speed up clinical examination