Specificity and sensitivity of transcranial sonography of the substantia nigra in the diagnosis of Parkinson's disease: prospective cohort study in 196 patients

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002613
Article Type:	Research
Date Submitted by the Author:	19-Jan-2013
Complete List of Authors:	Bouwmans, Angela; MUMC, Neurology Vlaar, Annemarie; St. Lucas Andreas Hospital, Neurology Mess, Werner; MUMC, Clin Neurophysiology Kessels, Alfons; MUMC, KEMTA Weber, Wim; MUMC, Neurology
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Radiology and imaging
Keywords:	Parkinson-s disease < NEUROLOGY, Ultrasound < RADIOLOGY & IMAGING, Neurophysiology < NEUROLOGY

Specificity and sensitivity of transcranial sonography of the substantia nigra in the diagnosis of Parkinson's disease: prospective cohort study in 196 patients

Angela E. P. Bouwmans MD [a], Annemarie M. M. Vlaar MD PhD[b], Werner H. Mess Prof MD PhD[c], Alfons Kessels MD MSc[d], and Wim E. J. Weber MD PhD [a]

Departments of [a]Neurology, of [c]Clinical Neurophysiology, and of [d]Clinical Epidemiology and Medical Technology Assesment (KEMTA), Maastricht University Medical Centre, Maastricht, The Netherlands

[b]Department of Neurology, Sint Lucas Andreas Hospital, Amsterdam, The Netherlands

Corresponding author: dr. W.E.J. Weber, Dept. of Neurology, Maastricht University Medical Centre, PO Box 5800, 6202 AZ Maastricht, The Netherlands

Tel.: +31-43/3875058

Fax. : +31-43/3877055

E-mail : wim.weber@mumc.nl

Keywords: transcranial sonography, substantia nigra, Parkinson's disease, diagnostic accuracy

Word count: 3989

Abstract Numerous ultrasound studies have suggested that a typical enlarged area of echogenicity in the substantia nigra (SN+) can help diagnose idiopathic Parkinson's disease (IPD). However almost all these studies were retrospective and involved patients with well-established diagnoses and long disease duration.

Objective Assessment of the diagnostic accuracy of transcranial sonography (TCS) of the substantia nigra in the patient with an undiagnosed parkinsonian syndrome of recent onset.

Design Prospective cohort study for diagnostic accuracy.

Setting Neurology outpatient clinics of two teaching hospitals in the Netherlands.

Patients 196 consecutive patients, who were referred to two neurology outpatient clinics for analysis of clinically unclear parkinsonism. Within two weeks of inclusion all patients also underwent a TCS and a ^{123}I-ioflupane Single Photon Emission Computer Tomography (FP-CIT SPECT) scan of the brain.

Outcome measures After two years, patients were re-examined by two movement disorder specialist neurologists for a final clinical diagnosis, that served as a surrogate gold standard for our study.

Results The final clinical diagnosis was IPD in 102 (52,0%) patients. Twenty-four (12,3%) patients were diagnosed with atypical parkinsonisms (APS) of which 8 (4,0%) multisystem atrophy (MSA), 6 (3,1%) progressive supranuclear palsy (PSP), 6 (3,1%) Lewy body dementia (LBD) and 4 (2,0%) corticobasal degeneration (CBD). Twenty-one (10,7%)

patients had a diagnosis of vascular parkinsonism (VP), 20 (10,2%) essential tremor (ET), 7 (3,6%) drug-induced parkinsonism (DIP) and 22 (11,2%) patients had no parkinsonism but a alternative diagnosis. The sensitivity of a SN+ for the diagnosis IPD was 0,40 (Confidence Interval (CI) 0,30-0,50) and the specificity 0,61 (CI 0,52-0,70). Hereby the positive predictive value (PPV) was 0,53 and the negative predictive value (NPV) 0,48. The sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was 0.88 (CI 0,81-0,95) and 0.68 (CI 0,58-0,76) with a PPV of 0,75 and a NPV of 0,84.

Conclusion The diagnostic accuracy of TCS in early stage Parkinson's disease is not sufficient for routine clinical use.

Clinicaltrials.gov identifier: NCT0036819

Article summary

Article focus

- We wanted to assess the diagnostic accuracy of transcranial sonography (TCS) of the substantia nigra in patients with an undiagnosed parkinsonian syndrome of recent onset.
- A large body of evidence suggests that TCS of the substantia nigra can help diagnose idiopathic Parkinson's disease (IPD). The problem is that almost all these studies were retrospective and involved patients with well-established diagnoses and long disease duration.

Key messages

- The diagnostic accuracy of TCS in early stage Parkinson's disease is not sufficient for routine clinical use.

Strengths and limitations of this study

Strength of our study is its guaranteed prospective nature: we registered this study prospectively and we carried it out exactly as proposed in the published protocol. It is the largest prospective study on this technique in this patient population up till now. At inclusion we excluded the patients with already a clear diagnosis, thus closely mimicking the clinical situation in which the neurologist would need an additional tool for diagnostic workup.

A limitation, as in all these studies, is the lack of an objective gold standard, i.e. neuropathological analysis. We used clinical diagnosis after 2 years follow-up as gold standard. Longer follow-up periods will probably increase diagnostic accuracy, but will also lead to higher attrition rates in these elderly populations.

Introduction

In clinical practice the diagnosis of idiopathic Parkinson's disease (IPD), delineating it from the atypical parkinsonisms (APS), vascular parkinsonism (VP), drug induced parkinsonism (DIP), and essential tremor (ET) is still difficult[1-8]. Especially in the early stage of these diseases a large group of patients is erroneously diagnosed, even by experienced movement disorder specialists, when one uses post-mortem findings as a gold standard[9-13]. Longer-term follow-up studies with clinical criteria as a gold standard found that IPD was frequently overdiagnosed initially[14, 15]. As these disorders demand vastly differing therapies along varying prognoses, a multitude of ancillary investigations has been proposed as aids in the early diagnosis of IPD[16-20]. Of all these, ^{123}I-ioflupane Single Photon Emission Computer Tomography (FP-CIT SPECT) scans are most widely used in routine clinical practice to diagnose IPD. But a substantial fraction of patients with early IPD have normal scans, and the costs and use of intravenous radio-active tracers are seen as important disadvantages of this technique[19].

The search for a cheaper and more patient-friendly technique to diagnose IPD has thus continued and over the last 10 years transcranial sonography (TCS) of the substantia nigra (SN) has emerged as a promising tool in this regard. Numerous ultrasound studies have found that a significant percentage of patients with IPD has a typical enlarged area of echogenicity in the substantia nigra (SN+), which is thought to be associated with increased iron concentrations [21-38]. Some of these studies have suggested that with this echofeature one can diagnose IPD with reasonable sensitivity and specificity. Further research along these lines found that TCS might also be used to delineate IPD from the APS [39-44], such as Multiple System Atrophy (MSA) and Progressive Supranuclear Palsy (PSP). These patients appear to have normal or only a moderately enhanced hyperechogenic SN as have patients with VP [45], ET [46-48] and DIP. Patients with Lewy Body dementia (LBD) [49] and

Cortical Basal Degeneration (CBD) [50] have been reported to share the same echofeature with IPD patients, and researchers have found that the accuracy of the differential diagnosis can be enhanced by additional assessments of the echogenicity of the basal ganglia. Hyperechogenicity of the lentiform nucleus is commonly seen in patients with CBD, whereas patients with IPD have this echofeature only rarely. Furthermore, research showed that the absence of bilateral marked SN+ discriminated IPD from LBD with a moderate to good sensitivity, and a good specificity and positive predictive value [49]. All these different findings combined could then give a ´diagnostic fingerprint´ for these disorders by following an algorithm we recently postulated [51].

However almost all studies were retrospective and involved patients with well-established diagnoses and long disease duration. These findings can thus not simply be extrapolated to the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. Up till now only one prospective study has assessed the diagnostic accuracy in patients with recent onset parkinsonian signs and symptoms [30]. This study was relatively small, excluded patients with tremor, and followed up patients for only 12 months.

We have now assessed the diagnostic accuracy of TCS of the SN in 196 patients referred by their general practitioner (GP) for analysis of a parkinsonian syndrome of recent onset. We used a clinical diagnosis after two years as a surrogate gold standard and also compared TCS with FP-CIT-SPECT scans.

Methods

Patients

This was a prospective study testing the diagnostic accuracy of TCS of the SN in patients who are referred by their GP for a first consultation by a neurologist because of recent-onset parkinsonism of unclear origin [52]. The Institutional Review Board (IRB) of the University Hospital Maastricht approved the study (MEC 05–228, 4 April 2006), and the study was registered prospectively under (ITRSCC) NCT0036819. The study protocol was published before the study started (weblink: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2034584) [52].

We considered 283 consecutive patients, who were referred to two neurology outpatient clinics for analysis of clinically unclear parkinsonism (Neurology Outpatient Clinic of the Maastricht University Medical Centre (MUMC) in Maastricht and the Orbis Medical centre in Sittard, The Netherlands). Patients, in whom a definite diagnosis could be made at the first visit, were excluded from the study (n=42). Hence, we enrolled 241 patients. After signing informed consent, upon entering the study, all subjects underwent a structured interview and a neurological examination (See Additional file 1 [52]). These tests were performed by a physician not treating the patient and blinded for information in the routine clinical records [52].

Within two weeks of inclusion all patients underwent a TCS of the SN, at the department of Neurophysiology of the two mentioned hospitals. In each hospital TCS was done by one specially trained investigator (P. Wuisman MD in Orbis Medical Centre, Sittard, and Prof. W. Mess (WHM) in the MUMC). WHM is a very experienced sonographer, who did additional training with Prof. D. Berg one of the pioneers of this technique [53]. To ensure validity of the TCS assessments among our two sonographers we had already done an interobserver

study, and found an acceptable interobserver agreement with kappa values in the 0.7- 0.8 range [54].

Patients in whom a TCS of the SN was not possible because of a non-accessible bone window were excluded from the study resulting eventually in a group of 196 patients. Within two weeks of inclusion all patients also underwent a FP-CIT SPECT scan of the brain as described in our protocol [52].

After two years, patients were re-examined by two movement disorder specialist neurologists for a final clinical diagnosis, that served as a surrogate gold standard for our study. The four consultant neurologists who alternatingly did these assessments were all specialists in movement disorders with more than ten years' experience in this field (Bert Anten MD PhD, Fred Vreeling MD PhD, Wim Weber MD PhD, and Ania Winogrodzka MD PhD). These investigators were blinded for all test results of these patients. In the planning of these visits we had made sure that neither one of the two neurologists had ever seen the patient. They were asked to interview and examine the patient, as they would normally do during a routine neurologic consultation. They were asked to fill out the same standard form as had been done by the including investigator during the first visit of the patient (see Additional file 1). Among other items this form contained the Unified Parkinson's Disease Rating Scale (UPDRS)-III score [55], and afterwards the neurologists received these scores of the patient at the first visit, so that they could evaluate whether the patient had had any progression on that scale. They also received the results of the brain scan, preferable a Magnetic Resonance Imaging (MRI) of the cerebrum however when not possible due to claustrophobia or devices not allowed in the MRI, a Computer Tomography (CT) of the brain. Each neurologist was asked then to reach a final clinical diagnosis of the parkinsonian syndrome using the diagnostic clinical criteria for IPD and APS [9, 56-59]. One investigator compared these scores and when there was no agreement, the two neurologists were asked to discuss these patients using their notes, in an

effort to reach agreement on the final diagnosis. In all cases except five patients, this discussion resulted in agreement on the final neurological diagnosis. Concerning the five patients to discussion, the diagnosis made at regular controls on the outpatient clinic of Neurology was taken as a third opinion and so a final diagnosis had been made.

TCS

One investigator per hospital and blinded to clinical information, did the ultrasounds using a SONOS 5500 (Philips, Eindhoven, The Netherlands). The examination took place in a darkened room with the patient already lying on the examination table before the investigator entered the room. This was in order to minimize the possible identification of a patient's clinical signs. Patient and investigator had been asked not to talk about medical information. TCS investigation was performed bilaterally through the pre-auricular bone window with a 2–4 MHz phased array transducer. The quality of the bone window was scored as good, moderate or inferior. Two different methods were applied for the evaluation of the SN. First, the presence or absence of an obviously visible SN was scored (qualitative method). Second, the area of an possible signal intensity was manually encircled and automatically calculated (quantitative method). This was only done when the increase of the hyperechogenicity was located in the anatomical distribution of the SN meaning showing a typically stripe-shaped configuration. Both the right and left SN were measured from both sides.

FP CIT SPECT

The SPECT scanning had been performed within 2 weeks of inclusion in the study. In this study FP-CIT (^{123}I-ioflupane, Nycomed, Amersham, U.K.) is used as presynaptic radiotracer. Medication (amphetamine, citalopram, fentanyl, fluoxetine, fluvoxamine, paroxetine, sertraline, venlafaxine) which could interfere with the radiotracer had been discontinued at

least 5 half life times. After intravenous injection of the tracer, SPECT measured baseline dopamine transporter integrity in the brain. SPECT was performed with a triple head camera (MultiSPECT3, Siemens, Ohio, USA) equipped with high-resolution collimators. A semi-automatic template model program was used to calculate the ratios between left striatal and right striatal and occipital regions respectively. Total time of acquisition was 30 minutes (45 seconds per frame for 40 views per detector). Zoom factor: 1.00 and the matrix size: 128 × 128. Filtered back-projection acquisition was performed. Images were filtered using a Butterworth filter with a cut-off value of: 0.4–0.5 and an order of 5. A division between the caudate nucleus and putamen was made. The ratios were corrected using Alderson's brain phantom, with known activities in the caudate nucleus and putamen. A binding of two standard deviations below healthy controls was considered as abnormal (FP-CIT 8.25 +-1.85 for putamen and 7.76 +-1.77 for caudate nucleus). Beside quantitative the scans were also judged visually by the same nuclear specialist blinded for the final clinical diagnosis. If quantitative and visual judgments did not agree the conclusion of visual judgment was taken.

Statistical analysis

Analyses were performed with SPSS, version 16.0. To determine the diagnostic performance of the SN+ and the FP-CIT SPECT we constructed Receiver Operating Characteristics (ROC) curves and calculated the Area under The Curve (AUC) and their p-values.

Role of the funding source

None.

Results

Patient characteristics

We had originally included 241 patients into the TCS study after approaching 283 possible candidates (See patients flowchart) in the period September 2006 until September 2008. The number of patients with no accessible temporal bone window was 45 (18,7%); these were slightly older (mean age of 72,4 versus 69,2 years) and there were by far more females (71% versus 26%) in this group compared to group of patients included in the study . This resulted in a group of 196 patients who had undergone an initial TCS. After two years 30 (15,3%) patients had died and 52 (26,5%) patients were not able or willing to undergo a second neurologic examination. The remaining 114 patients all underwent examination by two neurologists for a final clinical diagnosis of their movement disorder in the period September 2008 until September 2010. For the other 82 patients we derived a clinical diagnosis from the most recent clinical charts by the treating neurologist. To check the validity of this approach, we also derived these diagnoses from medical records for the 114 patients of whom we did have a gold standard diagnosis, and we found an agreement between these diagnoses with a kappa of 0.8. We also found no significant differences in the distribution of diagnoses between the patients groups with and without a gold standard diagnosis. But the group without the gold standard follow-up diagnosis did have a significantly higher age (70,5 versus 67,7 years, p= 0,034) and a higher UPDRS total score at inclusion (30,2 versus 22,8, p= 0,031).

The final clinical diagnosis was IPD in 102 (52,0%) patients. See for further division of the division of the final diagnoses table 1. The remaining 22 (11,2%) patients with no parkinsonism had alternative diagnoses like isolated tremor, orthostatic tremor, tardive dyskinesia, multi-infarction dementia, M. Alzheimer, stroke, hypoxic encephalopathy, and psychogenic disorders.

Table 1. Patient characteristics

	All patients (n=196)	IPD (n=102)	APS (n=24)	VP (n=21)	ET (n=20)	DIP (n=7)	No parkinsonism (n=22)
Mean age in years (SD)	69,2 (9,54)	68,5 (9,3)	69,6 (8,6)	76,3ζ (5,9)	69,4 (11,2)	63,1 (10,4)	67,2 (10,1)
Men in %	74,0%	71,6%	79,2%	85,7%	75,0%	85,7%	63,6%
Mean duration complaints in months (SD)	34,2 (43,57)	29,8 (41,7)	25,8 (20,8)	25,0 (22,4)	66,2† (56,1)	68,6 (61,1)	32,3 (52,6)
Mean score UPDRS-III at inclusion (SD)	13,7 (7,3)	13,2 (6,1)	17,8 (9,6)	17,7ζζ (8,6)	10,7†† (5,3)	14,9 (5,1)	9,9 (6,8)

IPD= Parkinson's disease, APS= atypical parkinsonian syndromes, VP= vascular parkinsonism, ET= essential tremor, DIP= drug induced parkinsonism, UPDRS-III = Unified Parkinson's Disease Rating Scale part III, ζ= significant higher age compared to ET, DIP and no parkinsonism, ζζ= significant higher UPDRS-III compared to IPD, ††= significant lower UPDRS-III compared to APS and VP, †= significant longer mean duration of complaints compared to IPD, APS and VP

Final diagnoses and SN

Table 2 gives the presence or absence of a SN+ related to the final diagnoses. The cut-off of 0.20 cm2 corresponds to the 75th percentile of hyperechogenic signal extent at the SN in a healthy population [21, 27, 39].

Table 2. Final diagnoses divided to the results of the transcranial sonography (TCS)

	IPD no.	MSA no.	PSP no.	LBD no.	CBD no.	VP no.	ET no.	DIP no.	No Parkinsonism no.	Total no.
No presence of hyperechogenic SN 0,20 cm2 or more	61	4	2	3	2	12	13	5	16	118
Presence of hyperechogenic SN 0,20 cm2 or more	41	4	4	3	2	9	7	2	6	78
Total	102	8	6	6	4	21	20	7	22	196

IPD= Parkinson's disease, MSA= multisystem atrophy, PSP= progressive supranuclear palsy, LBD= Lewy body dementia, VP= vascular parkinsonism, ET= essential tremor, DIP= drug induced parkinsonism

One can see that the presence and absence of the hyperechogenic SNs are distributed at random over the various diagnoses, without any preference for one particular diagnosis. We also found no significant difference for the maximum size or the sum of the area of the SN+ in the different diagnoses (see table 3, figure 1 and 2).

Table 3. Final diagnoses divided to the results of the FP CIT SPECT

	IPD no.	MSA no.	PSP no.	LBD no.	CBD no.	VP no.	ET no.	DIP no.	No Parkinsonism no.	Total no.
Normal	11	3	2	1	1	11	18	6	16	69
Abnormal	80	4	4	4	3	7	1	1	3	107
Total	91	7	6	5	4	18	19	7	19	176

IPD= Parkinson's disease, MSA= multisystem atrophy, PSP= progressive supranuclear palsy, LBD= Lewy body dementia, VP= vascular parkinsonism, ET= essential tremor, DIP= drug induced parkinsonism

The maximum size of the area of the SN+ is the one side of the mesencephalon on which the SN+ is the largest one. In a considerable amount of patients the SN+ is bilaterally present so the both areas of the SN+ are summed up resulting in the sum of the area. The mean area of the ROC curve is 0,541. The sensitivity of a SN+ for the diagnosis IPD was 0,40 (Confidence Interval (CI) 0,30-0,50) and the specificity 0,61 (CI 0,52-0,70). Hereby the positive predictive value (PPV) was 0,53 and the negative predictive value (NPV) 0,48. Because of earlier literature suggesting also a SN+ in the diagnoses LBD and CBS, we added these two groups to the IPD and came to the same sensitivity (0,41) and specificity (0,62).

Earlier research had suggested that the symmetry of the SN+ helps to differentiate between IPD and LBD, so we also analysed this feature. Of the 3 SN+ in the patients with LBD, 2 (67%) were bilaterally hyperintense. However, 29 (71%) of the 41 SN+ in the IPD patients, were bilaterally hyperintense, so in our population this echofeature had no diagnostic discriminatory value between the diagnoses LBD and IPD.

Final diagnoses and FP-CIT-SPECT scan results

176 patients also underwent a FPCIT-SPECT at initial work-up, around the same time when they underwent a TCD, see table 3. The sensitivity and specificity of FP-CIT SPECT scans

for diagnosing IPD was respectively 0.88 (CI 0,81-0,95) and 0.68 (0,58-0,76) with a PPV of 0.75 and a NPV of 0.84. Figure 3 shows the ROC Curve of FP-CIT-SPECT minimal uptake in the putamen and nucleus caudatus as a diagnostic performance to detect IPD. This was much better concerning the SPECT with a mean area of the ROC curve of 0,815 compared to the TCS. TCS findings were concordant with SPECT findings in 89 of 176 patients (p 0,36). In the 114 patients which had been re-examinated after a follow-up of two years, the SPECT scan and the TCS results were in agreement in only 50 patients (p 0,53).

Discussion

We have tried to assess the diagnostic accuracy of TCS in IPD, in the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. We thus assessed its accuracy in 241 consecutive patients referred by their GP for analysis of a parkinsonian syndrome of recent onset. We used a clinical diagnosis after two years as a surrogate gold standard and also compared TCS with FP-CIT-SPECT scans. We found no significant correlation between the SN+ and any of the final diagnoses in patients presenting with first symptoms of a hitherto undiagnosed parkinsonism. Sensitivity and specificity of SN+ for the diagnosis of IPD was 0.4 and 0.61 respectively. Hereby the positive predictive value (PPV) was 0.53 and the negative predictive value (NPV) 0.48. In contrast, we found that the sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was respectively 0.88 and 0.68 with a PPV of 0.75 and a NPV of 0.84.

Strength of our study is its guaranteed prospective nature: we registered this study prospectively and we carried it out exactly as proposed in the published protocol [52].

Another strength is its size: it is the largest prospective study on this technique in this patient population up till now. At inclusion we excluded the patients with already a clear diagnosis which has not been done in the study of Gaenslen. We have also tried to obtain the best possible surrogate gold standard clinical diagnosis. We did this by having our patients examined by a pair of independent experienced movement disorder specialists instead of using only the results of the imaging techniques done by Gaenslen. We also tried to increase its validity by observing a follow-up of two years. This appeared to be a relative maximum, as by that time already a substantial fraction of patients had passed away or had deteriorated in such a way that they did not want or were able to undergo another examination. However, the size of our study population is about three times as large compared to the group of Gaenslen. Apart from the real gold standard of the post-mortem examination, which seems less feasible in modern times, we think that this gold standard diagnosis of IPD is methodologically the highest achievable one. Implicitly, the follow-up is also a weakness, as it led to considerable attrition. We tried to circumvent this by deriving diagnoses from the medical charts of those patients who were not diagnosed by our pair of specialists. Although our validation experiment showed that there was good agreement between these two methods of obtaining final diagnoses, we cannot exclude that it may have biased our results. Simultaneous SPECT scans, which were reasonably accurate in diagnosing IPD in our study population, appear to confirm this relative lack of bias. In our population the FP-CIT SPECT scan did not reach a specificity of 100%, confirming an earlier report that a substantial fraction of early stage IPD patients have a normal SPECT scans [19].

We found substantially lower values for sensitivity and specificity of TCS to diagnose IPD than reported in earlier studies, including our own[20-49, [60]]. In diagnostic accuracy studies there are two major sources of variability: spectrum bias and test review bias [61]. Spectrum bias is the skewing of test parameters due to differences between study populations.

Test review bias is skewing of test parameters due to differences in the amount of clinical information available to the investigator interpreting the test result. We think that spectrum bias is the main cause of the substantial differences between ours and earlier studies. With one exception [30], all the earlier studies were retrospective and involved patients who had already been diagnosed clinically with definite IPD. These later-stage patients are obviously not the patients for whom one needs additional diagnostic tools such as a TCS, as these patients already have an unequivocal clinical diagnosis. Our study clearly show that results obtained in already diagnosed patients cannot be simply extrapolated to early stage, as yet undiagnosed patients.

Our results also differ from the only other prospective study [30]. We believe here spectrum bias also plays a role: Gaenslen et al excluded patients with resting tremor, which we did not. The establishment of a definite diagnosis also differed between our studies. Gaenslen et al were not able to reach a definite diagnosis in all patients, possibly due to the shorter follow-up (1 vs 2 years in our study).

We cannot rule out test review bias, as we did try to blind the TCS examiner, but not to great lengths. But this, if present, would have skewed the results of Gaenslen et al, and not ours, as we found less diagnostic accuracy in the TCS.

One could then argue that our TCS examiners were not experienced enough. Both our examiners had more than twenty years of experience in ultrasound, and one of us (WHM) spent considerable time, for this research project, training with Prof. Berg's group in Tübingen. As stated above, we had already done an inter-observer study, which yielded reasonable intra- and inter-rater reliability, in accordance with results by others [54, 62]. Results of TCS seem not be substantially influenced by the type of ultrasound device used [38], and we have in the past also found good diagnostic accuracy in later-stage IPD patients when studied retrospectively [60].

We thus feel that the crucial difference between earlier studies and ours is the prospective unselected nature of our patient population. Ours represented exactly the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. We show here that, in our hands, the TCS cannot be used reliably for that purpose.

Acknowledgements We would like to thank professor Daniela Berg, consultant neurologists Bert Anten, Fred Vreeling, Ania Winogrodzka and Marinus van Kroonenburgh for their helpful cooperation.

Authors' contributions AV, WW, WM, and AK had the idea and designed the protocol. Trial was done by AB, AV, WW, WM. Interpretation of the data was done by AB, WW, and AK. Paper was written by all authors.

Data sharing Dataset available from the corresponding author at Dryad repository, who will provide a permanent, citable and open access home for the dataset.

Conflict of Interest We have no conflicts of interest to declare.

Funding source This work was funded by the "Stichting Internationaal Parkinson Fonds, The Netherlands".

Legends to figures

Figure 1. Patiens flow chart

Figure 2. Boxplot of final diagnoses compared to the range of the maximum size of the hyperechogenic substantia nigra (SN+)

Figure 3. ROC Curve the maximum and the sum of the area of hyperechogenic substantia nigra (SN+) correlated with the final diagnosis Idiopathic Parkinson's Disease (IPD)

References

1. Kalra, S., D.G. Grosset, and H.T. Benamer, *Differentiating vascular parkinsonism from idiopathic Parkinson's disease: a systematic review.* Mov Disord, 2010. **25**(2): p. 149-56.
2. Diaz-Corrales, F.J., et al., *Clinical features and 123I-FP-CIT SPECT imaging in drug-induced parkinsonism and Parkinson's disease.* Eur J Nucl Med Mol Imaging, 2010. **37**(3): p. 556-64.
3. Wenning, G.K., et al., *What clinical features are most useful to distinguish definite multiple system atrophy from Parkinson's disease?* J Neurol Neurosurg Psychiatry, 2000. **68**(4): p. 434-40.
4. Osaki, Y., et al., *Accuracy of clinical diagnosis of progressive supranuclear palsy.* Mov Disord, 2004. **19**(2): p. 181-9.
5. Litvan, I., et al., *Accuracy of the clinical diagnosis of corticobasal degeneration: a clinicopathologic study.* Neurology, 1997. **48**(1): p. 119-25.
6. Wadia, P.M. and A.E. Lang, *The many faces of corticobasal degeneration.* Parkinsonism Relat Disord, 2007. **13 Suppl 3**: p. S336-40.
7. Litvan, I., et al., *Accuracy of the clinical diagnoses of Lewy body disease, Parkinson disease, and dementia with Lewy bodies: a clinicopathologic study.* Arch Neurol, 1998. **55**(7): p. 969-78.
8. Cardoso, F., S.T. Camargos, and G.A. Silva Junior, *Etiology of parkinsonism in a Brazilian movement disorders clinic.* Arq Neuropsiquiatr, 1998. **56**(2): p. 171-5.
9. Hughes, A.J., et al., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.* J Neurol Neurosurg Psychiatry, 1992. **55**(3): p. 181-4.
10. Munhoz, R.P., L.C. Werneck, and H.A. Teive, *The differential diagnoses of parkinsonism: findings from a cohort of 1528 patients and a 10 years comparison in tertiary movement disorders clinics.* Clin Neurol Neurosurg, 2010. **112**(5): p. 431-5.
11. Newman, E.J., et al., *Accuracy of Parkinson's disease diagnosis in 610 general practice patients in the West of Scotland.* Mov Disord, 2009. **24**(16): p. 2379-85.
12. Rajput, D.R., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease.* J Neurol Neurosurg Psychiatry, 1993. **56**(8): p. 938-9.
13. Ahlskog, J.E., *Diagnosis and differential diagnosis of Parkinson's disease and parkinsonism.* Parkinsonism Relat Disord, 2000. **7**(1): p. 63-70.
14. Caslake, R., et al., *Changes in diagnosis with follow-up in an incident cohort of patients with parkinsonism.* J Neurol Neurosurg Psychiatry, 2008. **79**(11): p. 1202-7.
15. Marshall, V.L., et al., *Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: a 3-year European multicenter study with repeat [123I]FP-CIT SPECT.* Mov Disord, 2009. **24**(4): p. 500-8.
16. McKinnon, J.H., et al., *Sniffing out Parkinson disease: can olfactory testing differentiate parkinsonian disorders?* Neurologist, 2007. **13**(6): p. 382-5.
17. Katzenschlager, R. and A.J. Lees, *Olfaction and Parkinson's syndromes: its role in differential diagnosis.* Curr Opin Neurol, 2004. **17**(4): p. 417-23.
18. Brooks, D.J., *Imaging approaches to Parkinson disease.* J Nucl Med, 2010. **51**(4): p. 596-609.
19. Vlaar, A.M., et al., *Meta-analysis of the literature on diagnostic accuracy of SPECT in parkinsonian syndromes.* BMC Neurol, 2007. **7**: p. 27.
20. Piccini, P. and A. Whone, *Functional brain imaging in the differential diagnosis of Parkinson's disease.* Lancet Neurol, 2004. **3**(5): p. 284-90.

21. Walter, U., et al., *Substantia nigra echogenicity is normal in non-extrapyramidal cerebral disorders but increased in Parkinson's disease.* J Neural Transm, 2002. **109**(2): p. 191-6.
22. Berg, D., C. Siefker, and G. Becker, *Echogenicity of the substantia nigra in Parkinson's disease and its relation to clinical findings.* J Neurol, 2001. **248**(8): p. 684-9.
23. Ressner, P., et al., *Hyperechogenicity of the substantia nigra in Parkinson's disease.* J Neuroimaging, 2007. **17**(2): p. 164-7.
24. Huang, Y.W., et al., *Transcranial imaging of substantia nigra hyperechogenicity in a Taiwanese cohort of Parkinson's disease.* Mov Disord, 2007. **22**(4): p. 550-5.
25. Spiegel, J., et al., *Transcranial sonography and [123I]FP-CIT SPECT disclose complementary aspects of Parkinson's disease.* Brain, 2006. **129**(Pt 5): p. 1188-93.
26. Becker, G., et al., *Degeneration of substantia nigra in chronic Parkinson's disease visualized by transcranial color-coded real-time sonography.* Neurology, 1995. **45**(1): p. 182-4.
27. Berg, D., et al., *Vulnerability of the nigrostriatal system as detected by transcranial ultrasound.* Neurology, 1999. **53**(5): p. 1026-31.
28. Berg, D., et al., *Echogenicity of the substantia nigra: association with increased iron content and marker for susceptibility to nigrostriatal injury.* Arch Neurol, 2002. **59**(6): p. 999-1005.
29. Kim, J.Y., et al., *Midbrain transcranial sonography in Korean patients with Parkinson's disease.* Mov Disord, 2007. **22**(13): p. 1922-6.
30. Gaenslen, A., et al., *The specificity and sensitivity of transcranial ultrasound in the differential diagnosis of Parkinson's disease: a prospective blinded study.* Lancet Neurol, 2008. **7**(5): p. 417-24.
31. Berg, D., et al., *Five-year follow-up study of hyperechogenicity of the substantia nigra in Parkinson's disease.* Mov Disord, 2005. **20**(3): p. 383-5.
32. Walter, U., et al., *Transcranial brain sonography findings in clinical subgroups of idiopathic Parkinson's disease.* Mov Disord, 2007. **22**(1): p. 48-54.
33. Walter, U., D. Skoloudik, and D. Berg, *Transcranial sonography findings related to non-motor features of Parkinson's disease.* J Neurol Sci, 2010. **289**(1-2): p. 123-7.
34. Schweitzer, K.J., et al., *Substantia nigra hyperechogenicity as a marker of predisposition and slower progression in Parkinson's disease.* Mov Disord, 2006. **21**(1): p. 94-8.
35. Weise, D., et al., *Substantia nigra echogenicity: A structural correlate of functional impairment of the dopaminergic striatal projection in Parkinson's disease.* Mov Disord, 2009. **24**(11): p. 1669-75.
36. Prestel, J., et al., *Predictive value of transcranial sonography in the diagnosis of Parkinson's disease.* Mov Disord, 2006. **21**(10): p. 1763-5.
37. Hagenah, J.M., et al., *Substantia nigra hyperechogenicity correlates with clinical status and number of Parkin mutated alleles.* J Neurol, 2007. **254**(10): p. 1407-13.
38. Mehnert, S., et al., *Transcranial sonography for diagnosis of Parkinson's disease.* BMC Neurol, 2010. **10**: p. 9.
39. Walter, U., et al., *Brain parenchyma sonography discriminates Parkinson's disease and atypical parkinsonian syndromes.* Neurology, 2003. **60**(1): p. 74-7.
40. Walter, U., et al., *Transcranial brain sonography findings in discriminating between parkinsonism and idiopathic Parkinson disease.* Arch Neurol, 2007. **64**(11): p. 1635-40.

41. Behnke, S., et al., *Differentiation of Parkinson's disease and atypical parkinsonian syndromes by transcranial ultrasound.* J Neurol Neurosurg Psychiatry, 2005. **76**(3): p. 423-5.
42. Okawa, M., et al., *Transcranial sonography of the substantia nigra in Japanese patients with Parkinson's disease or atypical parkinsonism: clinical potential and limitations.* Intern Med, 2007. **46**(18): p. 1527-31.
43. Ebentheuer, J., et al., *Substantia nigra echogenicity in progressive supranuclear palsy.* Mov Disord, 2010. **25**(6): p. 765-8.
44. Lauckaite, K., et al., *Specificity of transcranial sonography in parkinson spectrum disorders in comparison to degenerative cognitive syndromes.* BMC Neurol, 2012. **12**: p. 12.
45. Tsai, C.F., et al., *Transcranial color-coded sonography helps differentiation between idiopathic Parkinson's disease and vascular parkinsonism.* J Neurol, 2007. **254**(4): p. 501-7.
46. Doepp, F., et al., *Brain parenchyma sonography and 123I-FP-CIT SPECT in Parkinson's disease and essential tremor.* Mov Disord, 2008. **23**(3): p. 405-10.
47. Budisic, M., et al., *Distinguishing Parkinson's disease and essential tremor with transcranial sonography.* Acta Neurol Scand, 2009. **119**(1): p. 17-21.
48. Stockner, H., et al., *Midbrain sonography in patients with essential tremor.* Mov Disord, 2007. **22**(3): p. 414-7.
49. Walter, U., et al., *Sonographic discrimination of dementia with Lewy bodies and Parkinson's disease with dementia.* J Neurol, 2006. **253**(4): p. 448-54.
50. Walter, U., et al., *Sonographic discrimination of corticobasal degeneration vs progressive supranuclear palsy.* Neurology, 2004. **63**(3): p. 504-9.
51. Bouwmans, A.E., et al., *Transcranial sonography for the discrimination of idiopathic Parkinson's disease from the atypical parkinsonian syndromes.* Int Rev Neurobiol, 2010. **90**: p. 121-46.
52. Vlaar, A.M., et al., *Protocol of a prospective study on the diagnostic value of transcranial duplex scanning of the substantia nigra in patients with parkinsonian symptoms.* BMC Neurol, 2007. **7**: p. 28.
53. Berg, D., J. Godau, and U. Walter, *Transcranial sonography in movement disorders.* Lancet Neurol, 2008. **7**(11): p. 1044-55.
54. Vlaar, A., et al., *The Reliability of Transcranial Duplex Scanning in Parkinsonian Patients: Comparison of Different Observers and Ultrasound Systems.* Ultraschall Med, 2010.
55. *The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations.* Mov Disord, 2003. **18**(7): p. 738-50.
56. Gilman, S., et al., *Second consensus statement on the diagnosis of multiple system atrophy.* Neurology, 2008. **71**(9): p. 670-6.
57. Litvan, I., et al., *Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop.* Neurology, 1996. **47**(1): p. 1-9.
58. McKeith, I.G., et al., *Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium.* Neurology, 2005. **65**(12): p. 1863-72.
59. Mahapatra, R.K., et al., *Corticobasal degeneration.* Lancet Neurol, 2004. **3**(12): p. 736-43.
60. Vlaar, A.M., et al., *The predictive value of transcranial duplex sonography for the clinical diagnosis in undiagnosed parkinsonian syndromes: comparison with SPECT scans.* BMC Neurol, 2008. **8**: p. 42.

61. Bachmann, L.M., et al., *Multivariable adjustments counteract spectrum and test review bias in accuracy studies.* J Clin Epidemiol, 2009. **62**(4): p. 357-361 e2.
62. van de Loo, S., et al., *Reproducibility and diagnostic accuracy of substantia nigra sonography for the diagnosis of Parkinson's disease.* J Neurol Neurosurg Psychiatry, 2010. **81**(10): p. 1087-92.

The Lancet Neurology Reviewers' reports

"Transcranial sonography of the substantia nigra in the diagnosis of Parkinson's disease", by Bouwans et al

Authors'responses are in blue

Reviewer #1: statistician

In general this is a good study and well considered however the results at the moment do not seem to flow from all of the statements. The study is well designed with prospective follow-up. The figures are good descriptions of the data.

Major Points

1A. When the authors are describing the results you cannot just give the SD and confidence interval without the point estimate.

We assume the reviewer is referring to Table 1, where we give means of all data, and these are not all distributed normally, e.g. duration of complaints. We agree with the reviewer that giving the median in those cases would be formally correct, but the table will be more confusing to read, so we would like to keep it as it is**.**

1B. Also the results "significantly higher age compared to the patients with APS (SD 2,180, 95% Confidence Interval (CI) -12,230 up to 13,897)" even with the change to a decimal point these don't seem to be ages. Are they difference in ages? If they are difference in ages then how can the confidence interval so widely include 0 but be called significant?
There are quite a few problems like this.

The reviewer is correct, the minus-sign was incorrect and has been deleted.

1C. The authors must provide confidence intervals for the sensitivity and specificity values.

These are now provided.

1D. I am uncertain why you have a sensitivity and specificity for no parkinsonism as that should be the reference for the other values. What is it being compared to? Usually you have test +/- and then disease +/- for one sensitivity and specificity value.

We understand the confusion, and have removed Table 3.

1E. The word correlated should not really be used for these comparisons are they are not really correlations.

We have replaced this with "diagnostic performance to detect IPD".

1F. Were you able to make diagnoses for all the patients who didn't return to the study?

Yes, we did, as described on page 11.

1G. Are the results similar if you exclude this different sort of patient?

Yes, and this is now mentioned in the last paragraph of the Results section.

1H. There are a number of occasions when there are multiple tests between the different disease groups. There is then a real multiple testing problem. First overall differences should be shown before then looking into each individual group.

We understand the concern of this reviewer, but the goal of this study was not to show the differences between various diagnoses, but to assess the diagnostic accuracy of TCS in IPD. We thus prefer not to change this.

Minor Points
1I. There are a few non English "en" descriptions.

These have been corrected

Reviewer #2
In this prospective blinded study, the authors used transcranial sonography (TCS) and FP-CIT SPECT to study 196 patients with clinically uncertain parkinsonism, who were then followed by neurologists for two years until a final clinical diagnosis was obtained. The authors found that the sensitivity and specificity of SN+ for IPD was 0.40 and 0.61, respectively. These were in contrast to the sensitivity of 0.88 and the specificity 0.68 of FP-CIT SPECT for diagnosing early IPD. It was concluded that TCS cannot be used for early differential diagnosis of Parkinson's disease.
This is a potentially important manuscript. The study was performed rigorously according to a prospective blinded design. The authors explain clearly how the states of the individual subjects was determined at recruitment as well as the basis for the final clinical diagnosis at 2 years. These details are critical and will be much appreciated by the readership. Several issues, however, remain to be addressed:

2A. The authors have reported the sensitivity (0.40) and specificity (0.61) of SN+ for the diagnosis of IPD as the main results of the present study. In addition to these values, it is critical to calculate and report other clinically useful parameters such as positive predictive value (PPV) and negative predictive value (NPV). I calculated these parameters based on the data given in Table 2 and found that the PPV and NPV were 0.53 and 0.48 for TCS, respectively. The authors should verify these values and report them in the manuscript.

This is correct: recalculation shows a PPV of is 0,53 and a NPV of 0,48 als negative predictive value (NPV). We have now reported these in the paper: in Abstract, Results, and in Discussion section.

2B. The sensitivity (0.88) and specificity (0.68) of FPCIT SPECT for the PD diagnosis were reported in the Abstract and Discussion, but were not included in the Results. The authors

should report the sensitivity, specificity, PPV and NPV of FPCIT and provide the corresponding data in a manner similar to the TCS data presented in Table 2.

We have followed this suggestion, and given the data in the text and Table 2 (PPV 0,75; NPV 0,84)

2C. The authors may wish to report the findings in the 114 patients in whom a final clinical diagnosis was obtained by two specialists according to the gold standard. While the authors convey that these findings did not differ from those obtained in the whole group, more detail is necessary. Indeed, the results from the subgroup confirmed according to the gold standard will help to avoid the bias to which the authors refer in the manuscript (p. 14).

This is identical to the one raised under 1G. We have mentioned in the last paragraph of the Results section.

The authors discuss how their results differ from those reported in an earlier prospective TCS study (Gaenslen et al. Lancet Neurology, 2008). That group showed excellent differential diagnosis of IPD with TCS (sensitivity of 90.7%, specificity of 82.4%, and PPV of 92.9%). The present study, by contrast, suggests that TCS cannot be used for early differential diagnosis of IPD. Understanding the cause for this discrepancy seems critical. While the authors provide several potential reasons for this, several additional points should be discussed:

2D. Although Gaenslen et al. indicate that all 60 of their patients had parkinsonism of unclear cause at the time of enrollment, they also state that "a clinical diagnosis was made in 22 patients at baseline" (p. 419 of their paper). Based on this information, the Gaenslen et al. cohort appears to have been comprised of 38 patients with uncertain parkinsonism and 22 others who had a firm clinical diagnosis at baseline. This contrasts with the current study in which all the subjects initially had an unclear clinical diagnosis.

This is a good point, and we have added this observation to the start of the Discussion.

2E 12 of the 60 patients in the Gaenslen et al. study did not receive a definite clinical diagnosis at the conclusion of the study. In these subjects, the DAT or RAC images were used to provide the final clinical diagnosis - a highly questionable approach, considering that early on, the differential diagnostic information provided by these scans in not reliable. (Moreover, it is not reasonable to use the diagnosis provided by one type of non-definitive brain imaging method (DAT or RAC) to evaluate the accuracy of another scanning technique (TCS).

This consideration is now added to the Discussion.

2F Lastly, the sample size of the current study is three times that of the Gaenslen et al. study. Therefore, the computation of sensitivity and specificity is likely to be more accurate now. Given the importance of the topic, the authors may consider these issues and discuss them in their manuscript.

This consideration is now added to the Discussion.

2G The authors make clear that their findings are different from earlier studies, including their

own. Nonetheless, they do not explain why the TCS technique failed to differentiate early PD from other atypical syndromes. Is the signal unstable and less accurate at early disease stages? Is the signal-to-noise ratio too low? Is it technically more challenging to obtain a good signal in early patients? Is the target structure more difficult to delineate? Such a discussion would be very interesting.

These are all possibilities, but also speculative, and this is placing the burden of proof incorrectly. TCS looks promising in methodologically inferior studies, but fails to be accurate when studied in a clinically relevant setup.

Other points:
2H. It is surprising that of >200 patients enrolled there was no doubt regarding the final clinical diagnosis in virtually all cases after an average follow-up of 2 years - given particularly that most of the subjects were in an early disease stage. How many cases were there in which there was no initial agreement between the two raters regarding the final diagnosis? Moreover, the five subjects in whom there was no final agreement between the raters should have been excluded. Obviously, in these subjects there exists substantial uncertainty on the final diagnosis. Why was the third opinion viewed as reliable?

This is a good point, but clinical diagnosis is always arbitrary, and this is how we defined our gold standard, and we would not know a better method.

2I. In the methods the authors mention that the quality of bone window was rated (inferior/moderate/good). I assume that the quality of bone window has a considerable impact on the reliability of SN measurement. How did the authors account for this?

Patients with an inferior bone window were excluded, and we feel that, as this is an inherent problem with TCS, all the other patients should be included, as we would like to follow our "Intention-to-diagnose" analysis. When one starts excluding all kinds of results, one will end up with high accuracy numbers, but that would not be clinically meaningful.

2J. The authors should explain more clearly the qualitative and quantitative TCS methods. How were the authors able to perform the quantitative measurement in cases where there was in the qualitative assessment?

The quantitative value with no "obviously visible" SN is zero.

Minor points:
2K. No explanation is provided as to why 241 patients are mentioned in the title but only 196 subjects show up in the abstract. The reason for this needs to be made clear to the readership.

We have corrected the title (query identical to 3E)

2L. P.8 last paragraph: ...Medication (.. cocaine..), I wonder who prescribed this.

This was indeed “self-medication”, we have deleted this to avoid confusion.

Reviewer #3

This is an interesting study with similar aim but different result and conclusion to those published by TLN back in May 2008 by Gaenslen et al. It is impossible to review this paper without reading again (very carefully) the Gaenslen et al. paper. Here Bouwmans et al. found that the sensitivity and specificity of TCS to predict the diagnosis of idiopathic PD were much lower (40 and 61%) than those results shown by Gaenslen et al. (90 and 82.4%). Both studies enrolled patients referred to a movement disorder clinic with early parkinsonism of unknown origin and were studied at baseline with TCS and then followed prospectively over a certain period of time when a final clinical diagnoses was finally made, sometimes supported by functional neuroimaging of the DA system. Some methodological differences may account, in part, for the different results obtained between these two papers.

1. Bouwmans' sample size is larger (196 vs 60 individuals).
2. Bouwmans's observational period is longer (2 vs 1 year).
3. Bouwman's study included patients with tremoric parkinsonism and Gaenslen's not. This indicates that B. paper included subjects with essential tremor, an entity known to be associated with hyperechogenicity of the substabtia nigra (SN+) (Stockner et al. Movement Disorders 2007).
4. Bowmans study excluded those patients that a definite diagnosis (presumably PD in most of them) was made at the first visit (baseline) and Gaenslen's not.
5. Mean duration of parkinsonism at baseline was longer in B than in G study (34 vs 15 months).

Some comments

3A In methods section is not stated how the diagnoses of PD, DLB and the remaining diseases were established.

Diagnoses were reached according to recent guidelines; we have now added literature references for these.

3B In the methods section it is not stated how is defined the presence of hyperechogenicity (this is first known in the results section in page 11 and should stated in methods).

This is mentioned in the Methods section, on p. 8

3C In the methods section it is not stated how the putamen and caudate nucleus ratios are considered abnormal on DAT imaging. How the values of the healthy controls were obtained?

This is described on p. 9 in the Methods section.

3D In contrast to previous literature, the authors found TCS is a not a useful method to predict PD and to distinguish PD from other forms of parkinsonism when the patient is first referred at the movement disorder unit by his GP. The following unexpected results challenge previous knowledge in the field. Could the authors explain why 1) SN+ was only seen in 40%

of the patients with the final diagnosis of PD. This is the key point since the sensitivity in the current study of TCS to predict PD is much lower than previous literature that usually gives figures of 80-90%. 2) Also, the frequency of SN+ is surprisingly higher in subjects with no parkinsonism (6 out of 22) and vascular parkinsonism (9 out of 22). So I think that the discussion should be expanded and that 70% of the entire discussion should now be focused on three aspects: 1) the differences with the Gleansen study, 2) why TCS can also detects frequently SN+ in other entities than PD, and 3) why TCS only
detects 40% SN+ in PD when SN size does not increase over time in PD and previous literature found SN+ in 80-90 in PD.

Discussion is expanded (see also queries 2D-F) and we also raise the possibility that the SN+ is less stable than suggested by previous research, that focused mainly on patients with later stage IPD.

3E Title hast to be changed. place 196 subjects instead of 241, since 45 of them were not evaluated with TCs and were excluded.

This is corrected (identical to query 2K)

Reviewer #4

This is an interesting study. The results of the study are surprising referring both TCS and DAT-SPECT. The main reason given by the authors for their discrepant results to the literature is that they postulate that their study is the first prospective.

4A Several other studies in this field were prospective, too: e.g. Neurology. 2004 Aug 10;63(3):504-9. Sonographic discrimination of corticobasal degeneration vs progressive supranuclear palsy. Walter U, Dressler D, Wolters A, Probst T, Grossmann A, Benecke R.; e.g. Arch Neurol. 2007 Nov;64(11):1635-40. Transcranial brain sonography findings in discriminating between parkinsonism and idiopathic Parkinson disease. Walter U, Dressler D, Probst T, Wolters A, Abu-Mugheisib M, Wittstock M, Benecke R, e.g. Arch Neurol. 2011 Jul;68(7):932-7. Enlarged substantia nigra hyperechogenicity and risk for Parkinson disease: a 37-month 3-center study of 1847 older persons. Berg D, Seppi K, Behnke S, Liepelt I, Schweitzer K, Stockner H,
Wollenweber F, Gaenslen A, Mahlknecht P, Spiegel J, Godau J, Huber H, Srulijes K, Kiechl S, Bentele M, Gasperi A, Schubert T, Hiry T, Probst M, Schneider V, Klenk J, Sawires M, Willeit J, Maetzler W, Fassbender K, Gasser T, Poewe W., e.g.: Midbrain sonography in patients with essential tremor. Stockner H, Sojer M, K KS, Mueller J, Wenning GK, Schmidauer C, Poewe W. Mov Disord. 2007 Feb 15;22(3):414-7., e.g. Decreased striatal dopamine transporter uptake and substantia nigra hyperechogenicity as risk markers of synucleinopathy in patients with idiopathic rapid-eye-movement sleep behaviour disorder: a prospective study [corrected]. Iranzo A, Lomeña F, Stockner H, Valldeoriola F, Vilaseca I, Salamero M, Molinuevo JL, Serradell M, Duch J, Pavía J, Gallego J, Seppi K, Högl B, Tolosa E, Poewe W, Santamaria J; Sleep Innsbruck Barcelona (SINBAR) group. Lancet Neurol. 2010 Nov;9(11):1070-7. Epub 2010 Sep 16. Erratum in: Lancet Neurol. 2010 Nov;9(11):1045.

Ours was the first with an a priori hypothesis and a published protocol, and a follow-up period long enough to reach a reliable clinical diagnosis.

Major Comments:

4B diagnostic accuracy. Which criteria did the authors use for the different conditions. How did the authors exclude brain lesions as a cause for a DAT-deficit? How did the authors classify patients with VP without having performed structural brain imaging? How did the authors exclude the Parkinson Variant of PSP in patients with parkinsonism? Here structural brain imaging may help (MRI measurements of brainstem structures in patients with Richardson's syndrome, progressive supranuclear palsy-parkinsonism, and Parkinson's disease. Longoni G, Agosta F, Kostic VS, Stojkovic T, Pagani E, Stosic-Opincal T, Filippi M. Mov Disord. 2011 Feb 1;26(2):247-55. doi: 10.1002/mds.23293. Epub 2010 Dec 15.; MRI measurements predict PSP in unclassifiable parkinsonisms: a cohort study. Morelli M, Arabia G, Novellino F, Salsone M, Giofrè L, Condino F, Messina D, Quattrone A, Neurology. 2011 Sep 13;77(11):1042-7. Epub 2011 Aug 10.)

We have now added that the diagnosing neurologists had access to MRI and CT scans.

4C Please do not use the term CBD. It is now well-known and established in the movement disorders field, that the term cortico-basal syndrome CBS should be used if diagnosis is clinical, as about half of the patients presenting with CBS do have an alternate diagnosis(e.g.: The many faces of corticobasal degeneration. Wadia PM, Lang AE. Parkinsonism Relat Disord. 2007;13 Suppl 3:S336-40. Review.).

We respectfully disagree, as the above literature reference also uses the term CBD.

4D How was diagnosis of drug-induced parkinsonism DIP done? How did the authors exclude "unmasked DIP" (Eur J Nucl Med Mol Imaging. 2010 Mar;37(3):556-64. Epub 2009 Oct 28. Clinical features and 123I-FP-CIT SPECT imaging in drug-induced parkinsonism and Parkinson's disease. Diaz-Corrales FJ, Sanz-Viedma S, Garcia-Solis D, Escobar-Delgado T, Mir P.), a condition which seems to be common in elderly patients with DIP (Role of DAT-SPECT in the diagnostic work up of parkinsonism. Scherfler C, Schwarz J, Antonini A, Grosset D, Valldeoriola F, Marek K, Oertel W, Tolosa E, Lees AJ, Poewe W. Mov Disord. 2007 Jul 15;22(9):1229-38. Review.)?

On clinical grounds.

4E According to large scale prospective studies, DAT-SPECT should have a specificity near to 100% (e.g. Mov Disord. 2009 Mar 15;24(4):500-8. Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: a 3-year European multicenter study with repeat [123I]FP-CIT SPECT. Marshall VL, Reininger CB, Marquardt M, Patterson J, Hadley DM, Oertel WH, Benamer HT, Kemp P, Burn D, Tolosa E, Kulisevsky J, Cunha L, Costa D, Booij J, Tatsch K, Chaudhuri KR, Ulm G, Pogarell O, Höffken H, Gerstner A, Grosset DG.; e.g. (123I) beta-CIT and single-photon emission computed tomographic imaging vs clinical evaluation in Parkinsonian syndrome: unmasking an early diagnosis. Jennings DL, Seibyl JP, Oakes D, Eberly S, Murphy J, Marek K. Arch Neurol. 2004 Aug;61(8):1224-9.) This author is wondering on the low specificity of DAT-SPECT in this paper. As some of the clinical examiners have not their research field in movement disorder (searching the pubmed, it seems that 2 of the 4 clinical examiners are Movement Disorders Experts, one seems to be an expert in MS, another for dementia), this might explain the discrepancy to the available DAT-SPECT

literature. Another explanation might be, that the authors seem not to have excluded structural abnormalities with structural imaging for parkinsonism or DAT-deficits.

We do not think that publication history determines clinical expertise; we have explicitly stated what the experience of our 4 diagnosing neurologists is. Apart from that, this reviewer is mistaken that DAT-SPECT has 100% diagnostic specificity. We have added a literature reference to show this (ref 60 in the paper).

4F How did SPECT influence the diagnoses of the patients? How did the authors classify patients with ET or DIP having abnormal DAT-binding? Indeed DIP patients may develop PD (see above) and ET seems to be a risk factor for PD, especially if examining elderly people (Neuroepidemiology. 2011;37(1):1-10. Epub 2011 Jul 13. Association between essential tremor and other neurodegenerative diseases: what is the epidemiological evidence? LaRoia H, Louis ED.).

It did not.

4G How did the authors deal with the available evidence that patients with CBS and DLB seem to have hyperechogenitiy in the SN (Neurology. 2004 Aug 10;63(3):504-9. Sonographic discrimination of corticobasal degeneration vs progressive supranuclear palsy. Walter U, Dressler D, Wolters A, Probst T, Grossmann A, Benecke R. for review: Role of transcranial ultrasound in the diagnosis of movement disorders. Godau J, Berg D. Neuroimaging Clin N Am. 2010 Feb;20(1):87-101. Review.). Sensitivity analysis excluding these conditions or putting them together with PD should be performed.

We have done as the reviewer suggested, and found no significant changes: p=0,674 and sensitivity of 0,41 and specificity of 0,62. This is now added to the text.

4H Loss of 40% of the patients is another major comment. Please repeat analysis without these cases

Has been done, see queries 1G and 2C.

4I cut-off value of SN. This might differ between ultrasound machines. Therefore in multicentre studies, other authors have used cut-off value based on the SN echogenic area of their own controls (see Arch Neurol. 2011 Jul;68(7):932-7. Enlarged substantia nigra hyperechogenicity and risk for Parkinson disease: a 37-month 3-center study of 1847 older persons. Berg D, Seppi K, Behnke S, Liepelt I, Schweitzer K, Stockner H, Wollenweber F, Gaenslen A, Mahlknecht P, Spiegel J, Godau J, Huber H, Srulijes K, Kiechl S, Bentele M, Gasperi A, Schubert T, Hiry T, Probst M, Schneider V, Klenk J, Sawires M, Willeit J, Maetzler W, Fassbender K, Gasser T, Poewe W.). Indeed, a very recent study has found a cut-off value of 0.18 cm2 based on their data. (Mov Disord. 2012 Aug;27(9):1194-6. doi: 10.1002/mds.25071. Epub 2012 Jun 12. Is transcranial sonography useful to distinguish drug-induced parkinsonism from Parkinson's disease? Mahlknecht P, Stockner H, Kiechl S, Willeit J, Rastner V,
Gasperi A, Rungger G, Poewe W, Seppi K.) Therefore, the authors should calculate a cut-off value based on a control sample examined by the authors in their centres.

The official guideline on TCS technique by Daniela Berg does not include this, but a a value of 0.20 of 0.25 cm 2 is generally recommended. We refer to our own interobserver study should this reviewer question the reliability of our echographers (ref 54).

4J Table 3 should be changed. Sens/Spec/PPV/NPV should be referred to PD. A possibility would be to give the diagnostic accuracy data in that way: PD vs. all, PD vs. ET, PD vs.

As requested by another reviewer, this Table was removed.

4K Please give all TN/TP/FN/FP in an additional table. Maybe, it would be possible to merge tables 2 and 3.

For the sake of clarity we would like to refrain from this: reviewer 3 was able to calculate all PPVs and NPVs from table 2, and the relevant data for the SPECT scans are in Table 3.

4L Why did the authors did not give this information for DAT-SPECT?

We did: it is in Table 3..

4M The para on statistics is too short. One cannot understand, what the authors have done.

This is expanded, also in response to queries by reviewer 1.

4N Please shorten the para on patient characteristics in the results section. This para is hardly readable and includes redundant information, as many information in this para is given in the table. Try to avoid information given in the table.

We have changed this and now only the significant abnormalities are given in the Table with footnotes.

4M *What do the authors mean with "prospective unselected nature of our patient population".*

We could not find this sentence in our manuscript.

4O The English should be improved.

Has been done.

Reviewer #5:

MAJOR COMMENTS:

This is a well conducted, prospective study, particularly as the protocol was published before the study had even started. The TCS methodology is in agreement with standard procedures reported in the literature, and the study aims are highly relevant clinically -in that they aim to

delineate the ability of TCS to discriminate PD from atypical parkinsonian syndromes very early in their evolution, at time of first referral to a specialist movement disorders clinic, in contrast to many published studies where the clinical diagnosis is already well established without imaging.

5A A potential major criticism is that basal ganglia TCS and III ventricle TCS were not performed. A wide third ventricle and lentiform nucleus hyperechogenicity shown on TCS have been shown to be helpful in differentiating between iPD and atypical parkinsonism{Walter:2007ei}. These features raised sensitivity to atypical versus idiopathic PD from 72% (normal SN in atypicals) to 82% (normal SN and wide III ventricle and/or hyperechogenic lenticular nucleus in atypicals). The auuthors mention this paper in passing (p5, first sentence) but ignore the discriminatory value of those two features. This is potentially a major omission in a study claiming to have established lack of discriminatory potential of TCS in atypical parkinsonism.
- Admittedly other studies have not found that lenticular changes added to the discriminatory value of SN hyperechogenicity Gaenslen{Gaenslen:2008jf}. However, their reported SN sensitivity/specificity for differential diagnosis of typical/atypical PD was already very high (94.8%/90%) and so adding another discriminatory feature may not have been that efficient at this stage.

We would like to refrain from including these data. Reliability of echography of these structures is less established and our study was not set up to study this explicitly. We note that the other prospective study did not include these either (Gaenslen et al).

MINOR COMMENTS:
5B p11, paragraph 2, 3, 4; reporting differences between final diagnosis groups: comas used instead of dots for decimals;

This is now corrected.

5C no means provided for groups (present in the table), only SD;

This is not correct; they are given.

5D if confidence intervals refer to difference in age between VP and APS, etc, then the first one '-12,230 up to 13,897' is not significant; unless it's not a minus but a dash.

We have corrected this; identical query to 1B

Dr Heather Jane Brown
Senior Editor, The Lancet Neurology
32 Jamestown Rd, London NW1 7BY
UNITED KINGDOM
Tel: +44 (0)20 7424 4956
Fax: +44 (0)1865 853014
Email: heather.brown@lancet.com

BMC Neurology

BioMed Central

Study protocol

Open Access

Protocol of a prospective study on the diagnostic value of transcranial duplex scanning of the substantia nigra in patients with parkinsonian symptoms

Annemarie MM Vlaar[1], Angela EP Bouwmans[1], Marinus JPG van Kroonenburgh[2], Werner H Mess[3], Selma C Tromp[1,3], Piet GWM Wuisman[4], Alfons GH Kessels[5], Ania Winogrodzka[1] and Wim EJ Weber*[1]

Address: [1]Department of Neurology, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands, [2]Department of Nuclear Medicine University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands, [3]Department of Clinical Neurophysiology, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands, [4]Department of Clinical Neurophysiology, Maasland Hospital, P.O. Box 5500 6130 MB Sittard, The Netherlands and [5]Department of Clinical Epidemiology and Technology Assessment, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands

Email: Annemarie MM Vlaar - a.vlaar@neurologie.azm.nl; Angela EP Bouwmans - abouw@sneu.azm.nl; Marinus JPG van Kroonenburgh - mvkr@lnug.azm.nl; Werner H Mess - wme@fknf.azm.nl; Selma C Tromp - selma.tromp@fknf.azm.nl; Piet GWM Wuisman - p.wuisman@orbisconcern.nl; Alfons GH Kessels - akes@kemta.azm.nl; Ania Winogrodzka - hwin@sneu.azm.nl; Wim EJ Weber* - wweb@neurologie.azm.nl

* Corresponding author

Published: 4 September 2007

BMC Neurology 2007, **7**:28 doi:10.1186/1471-2377-7-28

Received: 11 July 2007

Accepted: 4 September 2007

This article is available from: http://www.biomedcentral.com/1471-2377/7/28

Abstract

Background: Parkinson's disease (PD) is the second most common neurodegenerative disorder. As there is no definitive diagnostic test, its diagnosis is based on clinical criteria. Recently transcranial duplex scanning (TCD) of the substantia nigra in the brainstem has been proposed as an instrument to diagnose PD. We and others have found that TCD scanning of substantia nigra duplex is a relatively accurate diagnostic instrument in patients with parkinsonian symptoms. However, all studies on TCD so far have involved well-defined, later-stage PD patients, which will obviously lead to an overestimate of the diagnostic accuracy of TCD.

We have therefore set out to conduct a prospective study testing the diagnostic accuracy of TCD in patients with a parkinsonism of unclear origin.

Methods/Design: We will enrol 250 consecutive patients, who are referred to neurology outpatient clinics of two teaching hospitals, for analysis of clinically unclear parkinsonism. Patients, whose parkinsonism is clearly diagnosable at the first visit, will be excluded from the study. All patients will undergo a TCD of the substantia nigra. As a surrogate gold standard we will use the consensus clinical diagnosis reached by two independent, blinded, movement disorder specialist neurologists after 2 years follow-up. At the time of TCD, patients will also undergo a SPECT scan of the brain.

Discussion: As this prospective trial enrols only patients with an early-stage parkinsonism, it will yield data on the diagnostic accuracy of TCD that is relevant to daily clinical practice: The neurologist needs a diagnostic tool that provides additional information in patients with a clinically indefinable parkinsonian syndrome. The above described observational longitudinal study was designed to explicitly study this aspect in the diagnostic process.

Trial registration: (ITRSCC) NCT00368199

Background

Parkinson's disease (PD) is the second most common neurodegenerative disorder with a life-time risk of 2 percent in males and 1.3 percent in females [1]. Diagnosis is based on clinical criteria. In most cases the diagnosis of PD is straightforward when cardinal clinical signs and symptoms as bradykinesia, rigidity, and resting tremor are present [2]. However, these main features of PD are shared, at least in part, by atypical parkinsonian syndromes (APS), like multi system atrophy (MSA), progressive supranuclear palsy (PSP) and corticobasal degeneration), essential tremor (ET), vascular parkinsonism (VP), drug induced parkinsonism (DIP), dementia with Lewy bodies and Alzheimer's disease. Besides delineating PD from the above parkinsonian syndromes, distinguishing PD from normality can also be difficult, especially in early stage of the disease [3].

The gold standard for the diagnosis of PD is post-mortem neuropathological examination. Clinicopathological studies show that 2–25 % of the patients with IPD are classified incorrectly in the final stage of their disease, even by specialists in movement disorders, with MSA and PSP accounting for most false positives [2,4,5]. Diagnostic accuracy is certainly less than 90% in earlier disease, as Litvan et al. found that the median sensitivity for the diagnosis of PD increased from 73% at the first visit to 80% to the last visit after a mean follow-up of 9 years, and the median positive predictive value increased from 46 to 64% [6].

A reliable test to diagnose PD is important for two reasons. Prognosis and medical treatment in the various parkinsonian syndromes differ considerably and an objective disease marker would facilitate the development of neuroprotective therapies [7,8]. Several procedures have been proposed to diagnose PD: functional imaging with Positron Emission Tomography (PET)-scan or Single Photon Emission Computer Tomography (SPECT), olfactory- and neuropsychological tests, and DNA tests [9-12].

At the moment neuro-imaging techniques like PET and SPECT are the most widely used diagnostic tools [8]. PET is at least as reliable as SPECT, but its use in routine clinical practice is limited by high costs and a relative short half-life of its radioactive tracers [13-16]. Despite its widespread use, there is no consensus about the value of SPECT scintigraphy in the differential diagnosis of PD [8,17-19]. The ability of SPECT scanning to discriminate PD from normality and other parkinsonian syndromes varies greatly among different studies. A major issue here is that many studies use well-defined later-stage patients that are obviously not representative for the diagnostic problem that one wants to solve with a SPECT.

A more recent addition to the diagnostic armamentarium of the neurologist is transcranial duplex scanning (TCD) of the substantia nigra (SN) in the brainstem. In 1994 Becker discovered that patients with PD had bilateral hyperechogenicity of the SN [12], probably caused by iron deposition [20,21]. Several publications confirmed this observation that up to 90% of PD patients have increased echo-intensity of the SN. In healthy subjects and in patients with ET or VP this hyperintensity of the SN is only found in 10–25% [20,22-29].

This technique has high inter-observer reliability [22,25,29]. In a pilot study with 45 patients with PD or APS who underwent SPECT and TCD we found a positive predictive value of 95% of an abnormal TCD for an abnormal FP-CIT SPECT scan [30].

Based on this study we hypothesized that TCD of substantia nigra is a tool deserving a place in the diagnostic work-up of PD/Parkinsonism patients. TCD is less costly and less invasive than SPECT [31]. Since a diagnostic test for parkinsonian syndromes is especially valuable in the early stage of disease(s), we devised a prospective diagnostic study with a clinical follow-up after 2 years as surrogate gold standard. As SPECT is currently the most widely used diagnostic tool in parkinsonian syndromes we included this in the study to directly compare the two techniques as to their diagnostic capacities in this field.

Methods/Design

Design

Observational, prospective, longitudinal study.

Setting

Consecutive patients will be recruited from the Neurology Outpatient Clinic of two hospitals: the University Hospital Maastricht in Maastricht and the Maasland Hospital in Sittard, The Netherlands. TCD will be done in the departments of Clinical Neurophysiology of the two above mentioned hospitals. SPECT scanning will be done in the departments of Nuclear Medicine of the two hospitals

Ethical approval

The Institutional Review Board (IRB) of the University Hospital Maastricht has approved the study (MEC 05–228, 4 April 2006). (This IRB also functions as IRB for the Maasland Hospital in Sittard, The Netherlands). All patients will be asked for informed consent through a standardised information form that is also approved by the Institutional Review Board.

Participants

250 consecutive patients with new parkinsonian signs and symptoms (of unclear origin at the time of visit) referred to the Neurology Outpatient Clinics of the University

Hospital Maastricht (n = 150) and the Maasland Hospital Sittard, (n = 100).

Inclusion criteria

1. Patients with parkinsonian signs and symptoms of unclear origin at the time of visit at the Neurology Outpatient Clinic. In his/her differential diagnosis the treating neurologist should be considering one of the following conditions: PD, MSA, and PSP. ET, VP or DIP.

2. Age older than 18 years.

Exclusion criteria

1. Patients presenting with a clear unequivocal diagnosis of their parkinsonism.

2. Patients whose life expectancy is less than the required follow-up of two years.

Methods

After informed consent all subjects will undergo a structured interview, neurological examination (See Additional file 1), TCD and SPECT within 6 weeks of the initial visit at the Outpatient Clinic. After two years all patients are re-examined by two movement disorder specialist neurologists for a final clinical diagnosis. This diagnosis serves as a surrogate gold standard to calculate the accuracy of SPECT and TCD to differentiate between PD and other types of parkinsonism.

A. Interview and neurological examination

After informed consent the patient is seen on the outpatient clinic for the inclusion interview and neurological examination by a third party physician (i.e. a physician not treating the patient, and blinded for information in the routine clinical records). In the structured interview a standard form with the following items are discussed: medical history, used drugs and effect, intoxications, duration of complaints, and most affected body- side (See attachment 1). The following clinimetric scales are scored: UPDRS (parts I, III and IV) [32], Hoehn and Yahr score [33], Hamilton Rating Scale [34] for depression and the SCOPA cognition scale [35]. The Sniffin Sticks smell test is done according to a standardised protocol [36,37]. Finally, the including physician will try to reach a probable diagnosis, strictly applying the UK Parkinson's Disease Society Brain Bank criteria [2].

B. SPECT

All subjects will undergo SPECT scanning within 6 weeks of inclusion in the study. In this study FP-CIT (^{123}I-ioflupane, Nycomed, Amersham, U.K.) is used as presynaptic radiotracer. Medication (amphetamine, citalopram, fentanyl, cocaine, fluoxetine, fluvoxamine, paroxetine, sertraline, venlafaxine) which could interfere with the radiotracer is discontinued at least 5 half life times. After intravenous injection of the tracer, SPECT measures baseline dopamine transporter integrity in the brain. SPECT is performed with a triple head camera (MultiSPECT3, Siemens, Ohio, USA) equipped with high-resolution collimators. A semi-automatic template model programme is used to calculate the ratios between left striatal and right striatal and occipital regions respectively. Total time of acquisition is 30 minutes (45 seconds per frame for 40 views per detector). Zoom factor: 1.00 and the matrix size: 128 × 128. Filtered back-projection acquisition is performed. Images are filtered using a Butterworth filter with a cut-off value of: 0.4–0.5 and an order of 5. A division between the caudate nucleus and putamen is made. The ratios are corrected using Alderson's brain phantom, with known activities in the caudate nucleus and putamen. A binding of two standard deviations below healthy controls is considered as abnormal (FP-CIT 8.25 +-1.85 for putamen and 7.76 +-1.77 for caudate nucleus). Beside quantitative the scans will be also judged visually by the same nuclear specialist blinded for the final clinical diagnosis. If quantitative and visual judgments do not agree the conclusion of visual judgment is taken (unpublished data).

C. Transcranial Duplex Scanning (TCD)

TCD investigation is performed bilaterally through the pre-auricular bone window with a 2–4 MHz phased array transducer (SONOS 5500; Philips, Eindhoven, the Netherlands) by an experienced sonographer, blinded for the clinical data and SPECT results. The quality of the temporal bone window, the SN and Raphe nuclei (RN) of all subjects are scored directly by the sonographer blinded for the final clinical diagnosis and SPECT result. The quality of the bone window is scored as good, moderate or inconclusive.

Two different methods are applied for the evaluation of the echointensity of the SN. Firstly, the presence or absence of an obviously visible bilateral hyperechogenic SN is scored (qualitative method). The SN are scored as hyperechointens, not hyperechointens or inconclusive (= no typical configuration of hyperechointensity or low quality of the temporal bone window). Secondly, the area of an eventually hyperechogenic SN will be measured quantitatively (quantitative method). Both the right and left SN are measured from both sides, i.e. both temporal bone windows. After encircling, the area is automatically calculated. A hyperechogenic area of at least 0.2 cm^2 is classified as characteristic for PD. The RN are scored as: invisible (= iso-intense), just visible, visible (= hyperintense) or inconclusive (= doubtful echointensity or low quality temporal bone window).

To determine inter-observer variability and to increase the power of the study a second sonographer will also judge the acquired echo data. A loop of 64 images of each patient will be acquired scanning the brainstem cranio-caudally and will be stored in order to allow for off-line analysis. Off-line the quality of the temporal bone window, SN and RN will scored by the second sonographer.

D. Regular Outpatient follow-up

The initial treating neurologist will remain responsible for the regular outpatient management of the patient included in the study. He or she will discuss the test results with the patient, and base his/her treatment plan on these. All further clinical decisions will be made by the treating neurologist.

E. Re-examination at two-year follow-up

Two years after inclusion, all patients will be re-examined separately by two independent movement disorder specialist neurologists blinded for the tests results. They will also be blinded for the clinical records of the treating neurologist. The same standard form as in the first visit is filled in (see Additional file 1) and they will be asked to reach a clinical diagnosis, independently from each other, according to generally accepted clinical criteria [38-44]. If these two diagnoses are not identical, the final diagnosis of this patient will be coded as inconclusive.

Data analysis

Our main hypothesis is that TCD is as sensitive as SPECT to differentiate PD from other parkinsonian disorders. For the power analyses we assumed a sensitivity of SPECT of 90%, based on the analysis of our own data on 248 consecutive patients [19]. Assuming this 90% for TCD sensitivity we can accept as lowest border of the 95% confidence interval, 86% or higher:

$SD = (\sqrt{}\ (p1\ (1\text{-}p1))/n \rightarrow p1 = 0.9$, sd 0.02, implying n = 190 patients, needed who have the hyperintensity of SN with TCD scanning.

In our pilot study 15% of the patients had an insufficient temporal bone window, so 224 patients are needed to compensate for the amount of inconclusive TCDs [28]. Based on this study we expect that 90% of all patients with inconclusive parkinsonism will ultimately have PD, so we set a target of initial 250 patients with unclear parkinsonism needed for this trial. We will calculate the sensitivity and specificity, positive predictive value, negative predictive value and diagnostic odd's ratio (OR) with its 95% confidence intervals (95% CI) of the first clinical judgment, TCD, FP-CIT SPECT scan and smell tests to predict the clinical diagnosis after 2 year follow-up. Accuracy is determined for all parkinsonian subgroups separately (PD versus APS, PD versus ET, and PD versus VP, PD versus DIP, and PD versus all other types of parkinsonism). For expected SPECT, TCD and smell tests scores for each parkinsonian disorder, see table 1.

Additionally we will determine the predictive value of TCD compatible with PD for an abnormal FP-CIT SPECT scan. Finally the inter-observer reliability for SN and RN judgement by TCD will be determined (Cohen's kappa test).

SPSS 11.0 for windows (SPSS, Chicago, IL, USA) and StataSE9 (Stata corporation, Texas, USA) will be used for statistical analysis.

Discussion

The hitherto published literature on TCD in parkinsonian syndromes are cross-sectional studies on clinically well-defined patient populations [20-30,45]. Although one has to start with these to study the diagnostic potential of a new technique, these kind of studies are obviously not

Table 1: Expected TCE, FP-CIT SPECT and odour recognition results for all parkinsonian disorders

Disease	FP-CIT SPECT abnormal, [19, 46]	abnormal SN TCD [47]	odour recognition deficit	cognition deficit
IPD	++	+++	+++	±
ET	Normal	normal	normal	normal
VP	normal *	normal	normal (?)	±
DIP **	normal (+)	normal (+)	normal (+)	normal (±)
MSA	++	±	+	±
PSP	++	± (?)	normal (±)	++

= normal FP-CIT tracer binding ratios
↓: FP-CIT binding of at least 2 sd. below healthy controls
* especially visual judgement together with CT or MRI
** At least 10% percent of the patients with DIP will develop to the PD [48]. So some patients in early stages of PD can theoretically present as DIP. Berg et al. investigated the relation between echointensity of SN on TCD and DIP after the start of antipsychotic drugs. Patients with serious parkinsonism scored higher echointensity as patients with mild or no parkinsonism [47].
*** In our retrospective trial 76% of the 27 patients with APS had FP-CIT binding lower as 2 standard deviations below healthy controls. In the 11 studies included in our meta-analysis this percentage varies form 67 to 100% [19, 46]

representative of the clinical problem that one wants to solve with a TCD in parkinsonian syndromes. The treating neurologist wants a diagnostic tool that provides additional information in patients with a clinically indefinable parkinsonian syndrome.

The above described observational longitudinal study was designed to explicitly study this aspect in the diagnostic process. We arbitrarily choose the clinical diagnosis after two years as the surrogate gold standard. This is, of course, not ideal, as there will always remain a small proportion of patients that is not definitely diagnosable after two years, and misdiagnoses (as opposed to the ultimate gold standard the post-mortem pathological analysis) are still possible. In an effort to tackle his last obstacle we require the final diagnosis to be shared by two independent, blinded, experienced movement disorders specialist neurologists.

We included SPECT scans in the study to enable us to make a direct comparison between SPECT and TCD as to their diagnostic accuracy. Although the use of SPECT scans in the diagnostic work-up of parkinsonian patients is still debated, it is widely used [8,19,46]. We feel that this will add to the clinical relevancy of our study results. Additionally, contributions of tests for smell, depression, cognition, in the diagnostic process can be assessed also.

Duration and expected study completion

4 years (2 years inclusion, 2 follow-up). Start recruiting 1-9-2006. Expected study completion date 1-10-2010.

Abbreviations

PD, Parkinson's disease; MSA, multiple system atrophy; PSP, progressive supranuclear palsy; VP, vascular parkinsonism; DIP, drug induced parkinsonism; ET, essential tremor; SPECT, Single Photon Emission Computer Tomography; PET, Positron Emission Tomography; FP-CIT, ^{123}I-ioflupane; TCD, transcranial duplex scanning; SN, substantia nigra; RN, Raphe nuclei; UPDRS, United Parkinson's Disease Rating Scale

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

All authors have read and approved the final manuscript. WM and WW initiated the study. AV, WW, ST, AK, and WW wrote the protocol. AV, AB, AW and WW will do the patient inclusion. AW will do the clinical examination after two years for a final diagnosis. AV, WW and AK will do the statistical calculations. AV, AB, AW, WW, WM, MK and ST will write the first draft of the paper.

Additional material

Additional file 1

Standard clinical scorings form. Standard form used to collect clinical data as described.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2377-7-28-S1.doc]

Acknowledgements

This study is funded by the "Stichting Internationaal Parkinson Fonds".

We thank Fred Vreeling (neurologist University Hospital Maastricht), Bert Anten, Fergus Rooyer (neurologists Maasland Hospital Sittard), Rosie Hustinx (clinical neurophysiology, Maasland Hospital Sittard), Hans van der Pol (nuclear medicine, Maasland hospital Sittard), Tanja Lustermans (PD nurse practitioner, Maasland Hospital Maastricht), and Mirella Waber (PD nurse practitioner, University Hospital Maastricht), for their cooperation.

References

1. Elbaz A, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA: **Risk tables for parkinsonism and Parkinson's disease.** *J Clin Epidemiol* 2002, **55:**25-31.
2. Hughes AJ, Daniel SE, Kilford L, Lees AJ: **Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.** *J Neurol Neurosurg Psychiatry* 1992, **55:**181-184.
3. Koller WC, Montgomery EB: **Issues in the early diagnosis of Parkinson's disease.** *Neurology* 1997, **49:**S10-25.
4. Watanabe H, Saito Y, Terao S, Ando T, Kachi T, Mukai E, Aiba I, Abe Y, Tamakoshi A, Doyu M, Hirayama M, Sobue G: **Progression and prognosis in multiple system atrophy: an analysis of 230 Japanese patients.** *Brain* 2002, **125:**1070-1083.
5. Rajput AH, Rozdilsky B, Rajput A: **Accuracy of clinical diagnosis in parkinsonism--a prospective study.** *Can J Neurol Sci* 1991, **18:**275-278.
6. Litvan I, MacIntyre A, Goetz CG, Wenning GK, Jellinger K, Verny M, Bartko JJ, Jankovic J, McKee A, Brandel JP, Chaudhuri KR, Lai EC, D'Olhaberriague L, Pearce RK, Agid Y: **Accuracy of the clinical diagnoses of Lewy body disease, Parkinson disease, and dementia with Lewy bodies: a clinicopathologic study.** *Arch Neurol* 1998, **55:**969-978.
7. Clarke CE: **Neuroprotection and pharmacotherapy for motor symptoms in Parkinson's disease.** *Lancet Neurol* 2004, **3:**466-474.
8. Piccini P, Whone A: **Functional brain imaging in the differential diagnosis of Parkinson's disease.** *Lancet Neurol* 2004, **3:**284-290.
9. Dujardin K, Defebvre L, Duhamel A, Lecouffe P, Rogelet P, Steinling M, Destee A: **Cognitive and SPECT characteristics predict progression of Parkinson's disease in newly diagnosed patients.** *J Neurol* 2004, **251:**1383-1392.
10. Ponsen MM, Stoffers D, Booij J, van Eck-Smit BLF, Wolters EC, Berendse HW: **Idiopathic hyposmia as a preclinical sign of Parkinson's disease.** *Ann Neurol* 2004, **56:**173-181.
11. Storch A, Hofer A, Kruger R, Schulz JB, Winkler J, Gerlach M: **New developments in diagnosis and treatment of Parkinson's disease--from basic science to clinical applications.** *J Neurol* 2004, **251 Suppl 6:**VI/33-8.
12. Becker G, Struck M, Bogdahn U, Becker T: **Echogenicity of the brainstem raphe in patients with major depression.** *Psychiatry Res* 1994, **55:**75-84.
13. Fernandez HH, Friedman JH, Fischman AJ, Noto RB, Lannon MC: **Is altropane SPECT more sensitive to fluoroDOPA PET for detecting early Parkinson's disease?** *Med Sci Monit* 2001, **7:**1339-1343.
14. Huang WS, Chiang YH, Lin JC, Chou YH, Cheng CY, Liu RS: **Cross-over study of (99m)Tc-TRODAT-1 SPECT and (18)F-FDOPA PET in Parkinson's disease patients.** *J Nucl Med* 2003, **44:**999-1005.

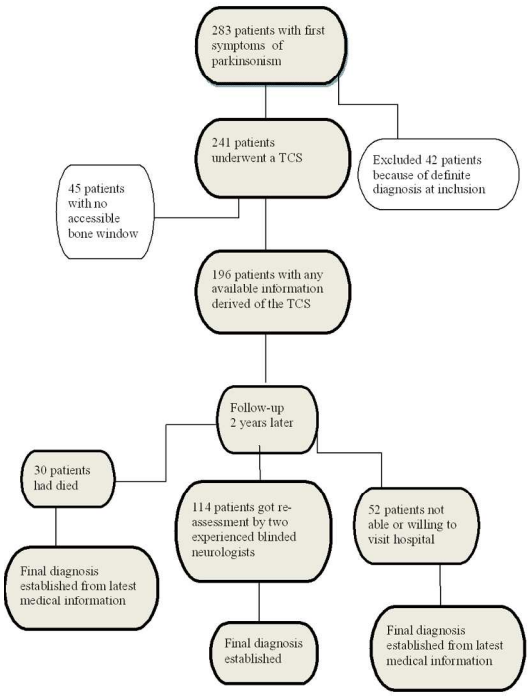
15. Ishikawa T, Dhawan V, Kazumata K, Chaly T, Mandel F, Neumeyer J, Margouleff C, Babchyck B, Zanzi I, Eidelberg D: **Comparative nigrostriatal dopaminergic imaging with iodine-123-beta CIT-FP/SPECT and fluorine-18-FDOPA/PET.** *J Nucl Med* 1996, **37:**1760-1765.
16. Schwarz J, Antonini A, Tatsch K, Kirsch CM, Oertel WH, Leenders KL: **Comparison of 123I-IBZM SPECT and 11C-raclopride PET findings in patients with parkinsonism.** *Nucl Med Commun* 1994, **15:**806-813.
17. Morrish PK: **How valid is dopamine transporter imaging as a surrogate marker in research trials in Parkinson's disease?** *Mov Disord* 2003, **18 Suppl 7:**S63-70.
18. Morrish P: **The meaning of negative DAT SPECT and F-Dopa PET scans in patients with clinical Parkinson's disease.** *Mov Disord* 2005, **20:**117; author reply 117-8.
19. Vlaar AMM, Kessels AGH, Nijs de T, Mess WH, Winogrodzka A, Kroonenburgh MJPG, Weber WEJ: **Additional value of SPECT imaging in comparison with clinical diagnosis in 248 patients with parkinsonism.** *Movement Disorders* 2007, **22:**S233.
20. Berg D, Roggendorf W, Schroder U, Klein R, Tatschner T, Benz P, Tucha O, Preier M, Lange KW, Reiners K, Gerlach M, Becker G: **Echogenicity of the substantia nigra: association with increased iron content and marker for susceptibility to nigrostriatal injury.** *Arch Neurol* 2002, **59:**999-1005.
21. Berg D, Grote C, Rausch WD, Maurer M, Wesemann W, Riederer P, Becker G: **Iron accumulation in the substantia nigra in rats visualized by ultrasound.** *Ultrasound Med Biol* 1999, **25:**901-904.
22. Berg D, Becker G, Zeiler B, Tucha O, Hofmann E, Preier M, Benz P, Jost W, Reiners K, Lange KW: **Vulnerability of the nigrostriatal system as detected by transcranial ultrasound.** *Neurology* 1999, **53:**1026-1031.
23. Becker G, Seufert J, Bogdahn U, Reichmann H, Reiners K: **Degeneration of substantia nigra in chronic Parkinson's disease visualized by transcranial color-coded real-time sonography.** *Neurology* 1995, **45:**182-184.
24. Walter U, Niehaus L, Probst T, Benecke R, Meyer BU, Dressler D: **Brain parenchyma sonography discriminates Parkinson's disease and atypical parkinsonian syndromes.** *Neurology* 2003, **60:**74-77.
25. Berg D, Siefker C, Becker G: **Echogenicity of the substantia nigra in Parkinson's disease and its relation to clinical findings.** *J Neurol* 2001, **248:**684-689.
26. Seppi K, Stockner S, Kiechl S, Schwaiger M, Sawires M, Willeit J, Poewe W: **Hyposmia, midbrain hyperechogenity and Parkinson's disease: findings in a population-based study.** *Movement Disorders* 2007, **22:**S10.
27. Bartova P, Skoloudik D, Fadma T, Ressner P, Kanovsky P, Herzig R: **Substantia nigra evaluation in atypical parkinsonian syndromes and Parkinson's disease.** *Movement Disorders* 2007, **22:**S142.
28. Tromp SC, Vlaar AMM, Kroonenburgh MJPG, Mess WH, Nijs de T, Winogrodzka A, Kessels AGH, Weber WEJ: **The diagnostic value of transcranial duplex scanning and SPECT imaging versus clinical diagnosis (clinical prospective study).** *Movement Disorders* 2007, **22:**S220.
29. Schweitzer KJ, Hilker R, Walter U, Burghaus L, Berg D: **Substantia nigra hyperechogenicity as a marker of predisposition and slower progression in Parkinson's disease.** *Mov Disord* 2006, **21:**94-98.
30. Tromp SC, Vlaar AMM, Weber WEJ, Mess WH: **Can ultrasonography of the substantia nigra replace FP-CIT-SPECT?(abstract).** *Parkinsonism Relat Disord* 2005, **11:**221.
31. Stockner H, Seppi K, Kiechl S, Schmidauer C, Schwaiger M, Sawires M, Willeit J, Poewe W: **Assessment of the feasibility of midbrain sonography in a population-based study.** *Movement Disorders* 2007, **22:**S146.
32. Fahn S, Elton RL: **The Unified Parkinson's Disease Rating Scale.** In *Recent developments in Parkinson's disease* Edited by: Fahn S, Marsden CD, Calne DB and Goldstein M. Florham Park, NJ, Marcmillan Health Care Information.; 1987:153-164.
33. Caplan B: **Neurological disorders. Course and treatment (handbook).** *Academic press San Diago, California* 1996.
34. Hamilton M: **A rating scale for depression.** *J Neurol Neurosurg Psychiatry* 1960, **23:**56-62.
35. Marinus J, Visser M, Verwey NA, Verhey FR, Middelkoop HA, Stiggelbout AM, van Hilten JJ: **Assessment of cognition in Parkinson's disease.** *Neurology* 2003, **61:**1222-1228.
36. Hummel T, Sekinger B, Wolf SR, Pauli E, Kobal G: **'Sniffin' sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold.** *Chem Senses* 1997, **22:**39-52.
37. Kobal G, Hummel T, Sekinger B, Barz S, Roscher S, Wolf S: **"Sniffin' sticks": screening of olfactory performance.** *Rhinology* 1996, **34:**222-226.
38. Litvan I, Agid Y, Jankovic J, Goetz C, Brandel JP, Lai EC, Wenning G, D'Olhaberriague L, Verny M, Chaudhuri KR, McKee A, Jellinger K, Bartko JJ, Mangone CA, Pearce RK: **Accuracy of clinical criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome).** *Neurology* 1996, **46:**922-930.
39. Daniel SE, Lees AJ: **Parkinson's Disease Society Brain Bank, London: overview and research.** *J Neural Transm Suppl* 1993, **39:**165-172.
40. Bain P, Brin M, Deuschl G, Elble R, Jankovic J, Findley L, Koller WC, Pahwa R: **Criteria for the diagnosis of essential tremor.** *Neurology* 2000, **54:**S7.
41. Zijlmans JC, Daniel SE, Hughes AJ, Revesz T, Lees AJ: **Clinicopathological investigation of vascular parkinsonism, including clinical criteria for diagnosis.** *Mov Disord* 2004, **19:**630-640.
42. Gilman S, Low PA, Quinn N, Albanese A, Ben-Shlomo Y, Fowler CJ, Kaufmann H, Klockgether T, Lang AE, Lantos PL, Litvan I, Mathias CJ, Oliver E, Robertson D, Schatz I, Wenning GK: **Consensus statement on the diagnosis of multiple system atrophy.** *J Neurol Sci* 1999, **163:**94-98.
43. McKeith IG, Galasko D, Kosaka K, Perry EK, Dickson DW, Hansen LA, Salmon DP, Lowe J, Mirra SS, Byrne EJ, Lennox G, Quinn NP, Edwardson JA, Ince PG, Bergeron C, Burns A, Miller BL, Lovestone S, Collerton D, Jansen EN, Ballard C, de Vos RA, Wilcock GK, Jellinger KA, Perry RH: **Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop.** *Neurology* 1996, **47:**1113-1124.
44. Wolters EC, Laar van T: **Bewegingsstoornissen.** In *VU Uitgeverij* Amsterdam, ; 2002.
45. Weise D, Lorenz R, Schliesser M, Reiners K, Classen J: **Transcranial sonography (TCS) of the substantia nigra (SN)-biological and clinical significance in patients with parkinsonian syndromes (PS).** *Movement Disorders* 2007, **22:**S194.
46. Vlaar AMM, Kroonenburgh MJPG, Kessels AGH, Tromp SC, Weber WEJ: **Diagnostic accuracy of SPECT in parkinsonian syndromes: a meta-analysis.** *Movement Disorders* 2007, **22:**S174.
47. Berg D, Jabs B, Merschdorf U, Beckmann H, Becker G: **Echogenicity of substantia nigra determined by transcranial ultrasound correlates with severity of parkinsonian symptoms induced by neuroleptic therapy.** *Biol Psychiatry* 2001, **50:**463-467.
48. Mena MA, de Yebenes JG: **Drug-induced parkinsonism.** *Expert Opin Drug Saf* 2006, **5:**759-771.

Pre-publication history

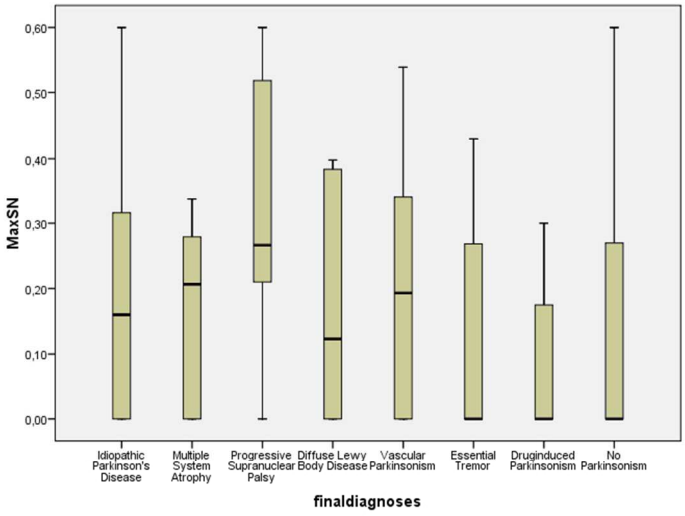
The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2377/7/28/prepub

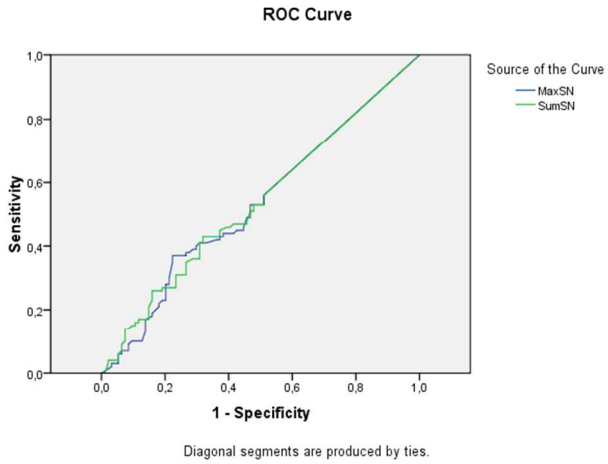
Flowchart. Trial profile:



Patiens flow chart
520x378mm (96 x 96 DPI)



Boxplot of final diagnoses compared to the range of the maximum size of the hyperechogenic substantia nigra (SN+)
254x142mm (96 x 96 DPI)



ROC Curve the maximum and the sum of the area of hyperechogenic substantia nigra (SN+) correlated with the final diagnosis Idiopathic Parkinson's Disease (IPD)
254x142mm (96 x 96 DPI)

STARD checklist for reporting of studies of diagnostic accuracy

(version January 2003)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	6
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	7
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	7
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	7
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	Prospective , see protcol
Test methods	7	The reference standard and its rationale.	9
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	9
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	9
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	9
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	9
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	10
	13	Methods for calculating test reproducibility, if done.	Ref. 54
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	7
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	7 and 12
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	Fig 1.
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	7-8
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	12
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	12-14
	20	Any adverse events from performing the index tests or the reference standard.	NA
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	Fig. 2
	22	How indeterminate results, missing data and outliers of the index tests were handled.	8-10
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	Ref. 54
	24	Estimates of test reproducibility, if done.	Ref.54
DISCUSSION	25	Discuss the clinical applicability of the study findings.	17-18

BMJ open

Specificity and sensitivity of transcranial sonography of the substantia nigra in the diagnosis of Parkinson's disease: prospective cohort study in 196 patients

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002613.R1
Article Type:	Research
Date Submitted by the Author:	02-Mar-2013
Complete List of Authors:	Bouwmans, Angela; MUMC, Neurology Vlaar, Annemarie; St. Lucas Andreas Hospital, Neurology Mess, Werner; MUMC, Clin Neurophysiology Kessels, Alfons; MUMC, KEMTA Weber, Wim; MUMC, Neurology
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Radiology and imaging
Keywords:	Parkinson-s disease < NEUROLOGY, Ultrasound < RADIOLOGY & IMAGING, Neurophysiology < NEUROLOGY

Specificity and sensitivity of transcranial sonography of the substantia nigra in the diagnosis of Parkinson's disease: prospective cohort study in 196 patients

Angela E. P. Bouwmans MD [a], Annemarie M. M. Vlaar MD PhD[b], Werner H. Mess Prof MD PhD[c], Alfons Kessels MD MSc[d], and Wim E. J. Weber MD PhD [a]

Departments of [a]Neurology, of [c]Clinical Neurophysiology, and of [d]Clinical Epidemiology and Medical Technology Assesment (KEMTA), Maastricht University Medical Centre, Maastricht, The Netherlands

[b]Department of Neurology, Sint Lucas Andreas Hospital, Amsterdam, The Netherlands

Corresponding author: dr. W.E.J. Weber, Dept. of Neurology, Maastricht University Medical Centre, PO Box 5800, 6202 AZ Maastricht, The Netherlands

Tel.: +31-43-3875058

Fax. : +31-43-3877055

E-mail : wim.weber@mumc.nl

Keywords: transcranial sonography, substantia nigra, Parkinson's disease, diagnostic accuracy

Word count: 3989

Abstract Numerous ultrasound studies have suggested that a typical enlarged area of echogenicity in the substantia nigra (SN+) can help diagnose idiopathic Parkinson's disease (IPD). Almost all these studies were retrospective and involved patients with well-established diagnoses and long disease duration.

Objective Assessment of the diagnostic accuracy of transcranial sonography (TCS) of the substantia nigra in the patient with an undiagnosed parkinsonian syndrome of recent onset.

Design Prospective cohort study for diagnostic accuracy.

Setting Neurology outpatient clinics of two teaching hospitals in the Netherlands.

Patients 196 consecutive patients, who were referred to two neurology outpatient clinics for analysis of clinically unclear parkinsonism. Within two weeks of inclusion all patients also underwent a TCS and a ^{123}I-ioflupane Single Photon Emission Computer Tomography (FP-CIT SPECT) scan of the brain (n=176).

Outcome measures After two years, patients were re-examined by two movement disorder specialist neurologists for a final clinical diagnosis, that served as a surrogate gold standard for our study.

Results Temporal acoustic windows were insufficient in 45 of 241 patients (18.67%). The final clinical diagnosis was IPD in 102 (52.0%) patients. Twenty-four (12.3%) patients were diagnosed with atypical parkinsonisms (APS) of which 8 (4.0%) multisystem atrophy (MSA), 6 (3.1%) progressive supranuclear palsy (PSP), 6 (3.1%) Lewy body dementia (LBD) and 4

(2.0%) corticobasal degeneration (CBD). Twenty-one (10.7%) patients had a diagnosis of vascular parkinsonism (VP), 20 (10.2%) essential tremor (ET), 7 (3.6%) drug-induced parkinsonism (DIP) and 22 (11.2%) patients had no parkinsonism but an alternative diagnosis. The sensitivity of a SN+ for the diagnosis IPD was 0.40 (Confidence Interval (CI) 0.30-0,50) and the specificity 0.61 (CI 0.52-0.70). Hereby the positive predictive value (PPV) was 0.53 and the negative predictive value (NPV) 0.48. The sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was 0.88 (CI 0.1-0.95) and 0.68 (CI 0.58-0.76) with a PPV of 0.75 and a NPV of 0.84.

Conclusion The diagnostic accuracy of TCS in early stage Parkinson's disease is not sufficient for routine clinical use.

Clinicaltrials.gov identifier: NCT0036819

Article summary

Article focus

- We wanted to assess the diagnostic accuracy of transcranial sonography (TCS) of the substantia nigra in patients with an undiagnosed parkinsonian syndrome of recent onset.
- A large body of evidence suggests that TCS of the substantia nigra can help diagnose idiopathic Parkinson's disease (IPD). The problem is that almost all these studies were retrospective and involved patients with well-established diagnoses and long disease duration.

Key messages

- The diagnostic accuracy of TCS in early stage Parkinson's disease is not sufficient for routine clinical use.

Strengths and limitations of this study

Strength of our study is its guaranteed prospective nature: we registered this study prospectively and we carried it out exactly as proposed in the published protocol. It is the largest prospective study on this technique in this patient population up till now. At inclusion we excluded the patients with already a clear diagnosis, thus closely mimicking the clinical situation in which the neurologist would need an additional tool for diagnostic workup.
A limitation, as in all these studies, is the lack of an objective gold standard, i.e. neuropathological analysis. We used clinical diagnosis after 2 years follow-up as gold standard. Longer follow-up periods will probably increase diagnostic accuracy, but will also lead to higher attrition rates in these elderly populations.

Introduction

In clinical practice the diagnosis of idiopathic Parkinson's disease (IPD), delineating it from the atypical parkinsonisms (APS), vascular parkinsonism (VP), drug induced parkinsonism (DIP), and essential tremor (ET) is still difficult[1-8]. Especially in the early stage of these diseases a large group of patients is erroneously diagnosed, even by experienced movement disorder specialists, when one uses post-mortem findings as a gold standard[9-13]. Longer-term follow-up studies with clinical criteria as a gold standard found that IPD was frequently overdiagnosed initially[14, 15]. As these disorders have varying prognoses, a multitude of ancillary investigations has been proposed as aids in the early diagnosis of IPD[16-20]. Of all these, ^{123}I-ioflupane Single Photon Emission Computer Tomography (FP-CIT SPECT) scans are most widely used in routine clinical practice to diagnose IPD. But a substantial fraction of patients with early IPD have normal scans, and the costs and use of intravenous radio-active tracers are seen as important disadvantages of this technique[19].

The search for a cheaper and more patient-friendly technique to diagnose IPD has thus continued and over the last 10 years transcranial sonography (TCS) of the substantia nigra (SN) has emerged as a promising tool in this regard. Numerous ultrasound studies have found that a significant percentage of patients with IPD has a typical enlarged area of echogenicity in the substantia nigra (SN+), which is thought to be associated with increased iron concentrations [21-38]. Some of these studies have suggested that with this echofeature one can diagnose IPD with reasonable sensitivity and specificity. Further research along these lines found that TCS might also be used to delineate IPD from the APS [39-44], such as Multiple System Atrophy (MSA) and Progressive Supranuclear Palsy (PSP). These patients appear to have normal or only a moderately enhanced hyperechogenic SN as have patients with VP [45], ET [46-48] and DIP. Patients with Lewy Body dementia (LBD) [49] and Cortical Basal Degeneration (CBD) [50] have been reported to share the same echofeature

with IPD patients, and researchers have found that the accuracy of the differential diagnosis can be enhanced by additional assessments of the echogenicity of the basal ganglia. Hyperechogenicity of the lentiform nucleus is commonly seen in patients with CBD, whereas patients with IPD have this echofeature only rarely. Furthermore, research showed that the absence of bilateral marked SN+ discriminated IPD from LBD with a moderate to good sensitivity, and a good specificity and positive predictive value [49]. All these different findings combined could then give a ´diagnostic fingerprint´ for these disorders by following an algorithm we recently postulated [51].

However almost all studies were retrospective and involved patients with well-established diagnoses and long disease duration. These findings can thus not simply be extrapolated to the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. Up till now only one prospective study has assessed the diagnostic accuracy in patients with recent onset parkinsonian signs and symptoms [30]. This study was relatively small, excluded patients with tremor, and followed up patients for only 12 months.

Methods

Patients

This was a prospective study testing the diagnostic accuracy of TCS of the SN in patients who are referred by their GP for a first consultation by a neurologist because of recent-onset parkinsonism of unclear origin [52]. The Institutional Review Board (IRB) of the University Hospital Maastricht approved the study (MEC 05–228, 4 April 2006), and the study was registered prospectively under (ITRSCC) NCT0036819. The study protocol was published before the study started [52].

We considered 283 consecutive patients, who were referred to two neurology outpatient clinics for analysis of clinically unclear parkinsonism (Neurology Outpatient Clinic of the Maastricht University Medical Centre (MUMC) in Maastricht and the Orbis Medical centre in Sittard, The Netherlands). Patients, in whom a definite diagnosis could be made at the first visit, were excluded from the study (n=42). Hence, we enrolled 241 patients. After signing informed consent, upon entering the study, all subjects underwent a structured interview and a neurological examination (See Additional file 1 [52]). These tests were performed by a physician not treating the patient and blinded for information in the routine clinical records [52].

Within two weeks of inclusion all patients underwent a TCS of the SN, at the department of Neurophysiology of the two mentioned hospitals. In each hospital TCS was done by one specially trained investigator (P. Wuisman MD in Orbis Medical Centre, Sittard, and Prof. W. Mess (WHM) in the MUMC). WHM is a very experienced sonographer, who did additional training with Prof. D. Berg one of the pioneers of this technique [53]. To ensure validity of the TCS assessments among our two sonographers we had already done an interobserver study, and found an acceptable interobserver agreement with kappa values in the 0.7- 0.8 range [54].

Patients in whom a TCS of the SN was not possible because of a non-accessible bone window were excluded from the study resulting eventually in a group of 196 patients. Within two weeks of inclusion all patients also underwent a FP-CIT SPECT scan of the brain as described in our protocol [52].

After two years, patients were re-examined by two movement disorder specialist neurologists for a final clinical diagnosis, that served as a surrogate gold standard for our study. The four consultant neurologists who alternatingly did these assessments were all specialists in movement disorders with more than ten years' experience in this field (Bert Anten MD PhD, Fred Vreeling MD PhD, Wim Weber MD PhD, and Ania Winogrodzka MD PhD). These investigators were blinded for all test results of these patients. In the planning of these visits we had made sure that neither one of the two neurologists had ever seen the patient. They were asked to interview and examine the patient, as they would normally do during a routine neurologic consultation. They were asked to fill out the same standard form as had been done by the including investigator during the first visit of the patient (see Additional file 1). Among other items this form contained the Unified Parkinson's Disease Rating Scale (UPDRS)-III score [55], and afterwards the neurologists received these scores of the patient at the first visit, so that they could evaluate whether the patient had had any progression on that scale. They also received the results of the brain scan, preferable a Magnetic Resonance Imaging (MRI) of the cerebrum however when not possible due to claustrophobia or devices not allowed in the MRI, a Computer Tomography (CT) of the brain. Each neurologist was asked then to reach a final clinical diagnosis of the parkinsonian syndrome using the diagnostic clinical criteria for IPD and APS [9, 56-59]. One investigator compared these scores and when there was no agreement, the two neurologists were asked to discuss these patients using their notes, in an effort to reach agreement on the final diagnosis. In all cases except five patients, this discussion resulted in agreement on the final neurological diagnosis. Concerning the five

patients to discussion, the diagnosis made at regular controls on the outpatient clinic of Neurology was taken as a third opinion and so a final diagnosis had been made.

TCS

One investigator per hospital and blinded to clinical information, did the ultrasound imaging (sonography) with a SONOS 5500 (Philips, Eindhoven, The Netherlands). The examination took place in a darkened room with the patient already lying on the examination table before the investigator entered the room. This was done to minimize the possible identification of a patient's clinical signs. Patient and investigator had been asked not to talk about medical information.

TCS investigation was performed bilaterally through the pre-auricular bone window with a 2–4 MHz phased array transducer. The quality of the bone window was scored as good, moderate or inferior. Two different methods were applied for the evaluation of the SN. First, the presence or absence of an obviously visible SN was scored (qualitative method). Second, the area of an possible signal intensity was manually encircled and automatically calculated (quantitative method). This was only done when the increase of the hyperechogenicity was located in the anatomical distribution of the SN meaning showing a typically stripe-shaped configuration. Both the right and left SN were measured from both sides.

FP CIT SPECT

The SPECT scanning was done within 2 weeks of inclusion in the study. In this study FP-CIT (^{123}I-ioflupane, Nycomed, Amersham, U.K.) was used as presynaptic radiotracer. Medication (amphetamine, citalopram, fentanyl, fluoxetine, fluvoxamine, paroxetine, sertraline, venlafaxine) which could interfere with the radiotracer had been discontinued at least 5 half life times. After intravenous injection of the tracer, SPECT measured baseline dopamine

transporter integrity in the brain. SPECT was performed with a triple head camera (MultiSPECT3, Siemens, Ohio, USA) equipped with high-resolution collimators. A semi-automatic template model program was used to calculate the ratios between left striatal and right striatal and occipital regions respectively. Total time of acquisition was 30 minutes (45 seconds per frame for 40 views per detector). Zoom factor: 1.00 and the matrix size: 128 × 128. Filtered back-projection acquisition was performed. Images were filtered using a Butterworth filter with a cut-off value of: 0.4–0.5 and an order of 5. A division between the caudate nucleus and putamen was made. The ratios were corrected using Alderson's brain phantom, with known activities in the caudate nucleus and putamen. A binding of two standard deviations below healthy controls was considered as abnormal (FP-CIT 8.25 +-1.85 for putamen and 7.76 +-1.77 for caudate nucleus). Beside quantitative the scans were also judged visually by the same nuclear specialist blinded for the final clinical diagnosis. If quantitative and visual judgments did not agree the conclusion of visual judgment was taken.

Statistical analysis

Analyses were performed with SPSS, version 16.0. To determine the diagnostic performance of the SN+ and the FP-CIT SPECT we constructed Receiver Operating Characteristics (ROC) curves and calculated the Area under The Curve (AUC) and their p-values.

Role of the funding source

None.

Results

Patient characteristics

We had originally included 241 patients into the TCS study after approaching 283 possible candidates (See patients flowchart) in the period September 2006 until September 2008. The number of patients with no accessible temporal bone window was 45 (18.7%); these were slightly older (mean age of 72.4 versus 69.2 years) and there were by far more females (71% versus 26%) in this group compared to group of patients included in the study . This resulted in a group of 196 patients who had undergone an initial TCS. After two years 30 (15.3%) patients had died and 52 (26.5%) patients were not able or willing to undergo a second neurologic examination. The remaining 114 patients all underwent examination by two neurologists for a final clinical diagnosis of their movement disorder in the period September 2008 until September 2010. For the other 82 patients we derived a clinical diagnosis from the most recent clinical charts by the treating neurologist. To check the validity of this approach, we also derived these diagnoses from medical records for the 114 patients of whom we did have a gold standard diagnosis, and we found an agreement between these diagnoses with a kappa of 0.8. We also found no significant differences in the distribution of diagnoses between the patients groups with and without a gold standard diagnosis. But the group without the gold standard follow-up diagnosis did have a significantly higher age (70.5 versus 67.7 years, p= 0.034) and a higher UPDRS total score at inclusion (30.2 versus 22.8, p= 0.031).

The final clinical diagnosis was IPD in 102 (52.0%) patients. See for further division of the division of the final diagnoses table 1. The remaining 22 (11.2%) patients with no parkinsonism had alternative diagnoses like isolated tremor, orthostatic tremor, tardive dyskinesia, multi-infarction dementia, Alzheimer's disease, stroke, hypoxic encephalopathy, and psychogenic disorders.

Table 1. Patient characteristics

	All patients (n=196)	IPD (n=102)	APS (n=24)	VP (n=21)	ET (n=20)	DIP (n=7)	No parkinsonism (n=22)
Mean age in years (SD)	69.2 (9.54)	68.5 (9.3)	69.6 (8.6)	76.3ζ (5.9)	69.4 (11.2)	63.1 (10.4)	67.2 (10.1)
Men in %	74.0%	71.6%	79.2%	85.7%	75.0%	85.7%	63.6%
Mean duration complaints in months (SD)	34.2 (43.57)	29.8 (41.7)	25.8 (20.8)	25.0 (22.4)	66.2† (56.1)	68.6 (61.1)	32.3 (52.6)
Mean score UPDRS-III at inclusion (SD)	13.7 (7.3)	13.2 (6.1)	17.8 (9.6)	17.7ζζ (8.6)	10.7†† (5.3)	14.9 (5.1)	9.9 (6.8)

IPD = Parkinson's disease, APS = atypical parkinsonian syndromes, VP = vascular parkinsonism, ET = essential tremor, DIP = drug induced parkinsonism, UPDRS-III = Unified Parkinson's Disease Rating Scale part III, ζ = significant (p value below 0.05) higher age compared to ET, DIP and no parkinsonism, ζζ = significant (p value below 0.05) higher UPDRS-III compared to IPD, †† = significant (p value below 0.05) lower UPDRS-III compared to APS and VP, † = significant (p value below 0.05) longer mean duration of complaints compared to IPD, APS and VP

Final diagnoses and SN

Table 2 gives the presence or absence of a SN+ related to the final diagnoses. The cut-off of 0.20 cm2 corresponds to the 75th percentile of hyperechogenic signal extent at the SN in a healthy population [21, 27, 39].

Table 2. Final diagnoses divided to the results of the transcranial sonography (TCS)

SN	IPD	MSA	PSP	LBD	CBD	VP	ET	DIP	No Parkinsonism	Total
	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.
SN- (%)	61 (60)	4 (50)	2 (33)	3 (50)	2 (50)	12 (57)	13 (65)	5 (71)	16 (73)	118 (60)
SN+ (%)	41 (40)	4 (50)	4 (67)	3 (50)	2 (50)	9 (43)	7 (35)	2 (29)	6 (27)	78 (40)
Total	102	8	6	6	4	21	20	7	22	196

SN= substantia nigra, SN- = no presence of hyperechogenic SN 0,20 cm2 or more, SN+ = presence of hyperechogenic SN 0,20 cm2 or more , IPD = Parkinson's disease, MSA = multisystem atrophy, PSP = progressive supranuclear palsy, LBD = Lewy body dementia, VP = vascular parkinsonism, ET = essential tremor, DIP = drug induced parkinsonism

One can see that the presence and absence of the hyperechogenic SNs are distributed at random over the various diagnoses, without any preference for one particular diagnosis. We also found no significant difference for the maximum size or the sum of the area of the SN+ in the different diagnoses (see table 3, figure 1 and 2). The maximum size of the area of the SN+ is the one side of the mesencephalon on which the SN+ is the largest one. In a considerable number of patients the SN+ was bilaterally present, and then both areas of the SN+ were added. We were not able to obtain better diagnostic discrimination with other TCS cut-offs. E.g., when we lowered the sensitivity threshold to an absolute minimum of 0.7, we obtained a cut-off of 0,3 cm2, but then specificity was 0.29.

The mean area of the ROC curve was 0.541. The sensitivity of a SN+ for the diagnosis IPD was 0.40 (Confidence Interval (CI) 0.30-0.50) and the specificity 0.61 (CI 0.52-0.70). Positive predictive value (PPV) was 0.53 and the negative predictive value (NPV) 0.48. Because earlier research suggested that SN+ can help diagnose LBD and CBS, we added these two groups to the IPD and recalculated, reaching the same sensitivity (0.41) and specificity (0.62). Earlier research had suggested that symmetry of the SN+ helps to differentiate between IPD and LBD. Of the 3 SN+ in the patients with LBD, 2 (67%) were bilaterally hyperintense. However, 29 (71%) of the 41 SN+ in the IPD patients, were bilaterally hyperintense, so in our population this echofeature had no diagnostic discriminatory value between the diagnoses LBD and IPD. As the current view is that these diagnoses clinically overlap, it may be inappropriate to consider the diagnosis of DLB instead of PD to be a diagnostic error.

Final diagnoses and FP-CIT-SPECT scan results

176 patients also underwent a FPCIT-SPECT at initial work-up, around the same time when they underwent a TCD (Table 3). The sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was respectively 0.88 (CI 0.81-0.95) and 0.68 (0.58-0.76) with a PPV of 0.75 and a NPV of 0.84. Figure 3 shows the ROC Curve of FP-CIT-SPECT minimal uptake in the putamen and nucleus caudatus as a diagnostic performance to detect IPD. TCS findings were concordant with SPECT findings in 89 of 176 patients (p= 0.36).

Table 3. Final diagnoses divided to the results of the FP CIT SPECT

FP CIT SPECT	IPD no.	MSA no.	PSP no.	LBD no.	CBD no.	VP no.	ET no.	DIP no.	No Parkinsonism no.	Total no.
Normal (%)	11 (12)	3 (43)	2 (33)	1 (20)	1 (25)	11 (61)	18 (95)	6 (86)	16 (84)	69 (39)
Abnormal (%)	80 (88)	4 (57)	4 (67)	4 (80)	3 (75)	7 (39)	1 (5)	1 (14)	3 (16)	107 (61)
Total	91	7	6	5	4	18	19	7	19	176

IPD= Parkinson's disease, MSA= multisystem atrophy, PSP= progressive supranuclear palsy, LBD= Lewy body dementia, VP= vascular parkinsonism, ET= essential tremor, DIP= drug induced parkinsonismIn the 114 patients which had been re-examined after a follow-up of two years, the SPECT scan and the TCS results were in agreement in only 50 patients (p= 0.53).

This concordance of TCS and SPECT data was randomly distributed over the diagnostic groups. We also studied diagnostic accuracy in terms of delineating PD from non-parkinsonian (APS) syndromes. When we grouped all IPD diagnoses together with all APS diagnoses vs. the rest, i.e. ET, DIP, VP, etc., we found similar specificity and sensitivity of TCS, respectively 0.67 and 0.43. In this analysis sensitivity of SPECT remained 0.84 and specificity increased to 0.84.

Discussion

We have tried to assess the diagnostic accuracy of TCS in IPD, in the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. We thus assessed its accuracy in 241 consecutive patients referred by their GP for analysis of a parkinsonian syndrome of recent onset. We used a clinical diagnosis after two years as a surrogate gold standard and

also compared TCS with FP-CIT-SPECT scans. Sensitivity and specificity of SN+ for the diagnosis of IPD was 0.4 and 0.61 respectively. Positive predictive value (PPV) was 0.53 and the negative predictive value (NPV) 0.48. In contrast, we found that the sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was respectively 0.88 and 0.68 with a PPV of 0.75 and a NPV of 0.84. In our hands, temporal acoustic windows were insufficient in a relatively high proportion of patients: 18.67%.

Strength of our study is its guaranteed prospective nature: we registered this study prospectively and we carried it out exactly as proposed in the published protocol [52]. Another strength is its size: it is the largest prospective study on this technique in this patient population up till now. At inclusion we excluded the patients with already a clear diagnosis. We have also tried to obtain the best possible surrogate gold standard clinical diagnosis. We did this by having our patients examined by a pair of independent experienced movement disorder specialists. The accepted gold standard is postmortem neuropathological examination, but this is hardly feasible anymore in modern times, as relatives are reluctant to give permission for this. So, the methodologically highest achievable gold standard is clinical examination after several years. This follow-up is essential as the diagnostic criteria contain several items that can only be assessed after a certain amount of time (levodopa response, progression, other diagnoses. The follow-up of two years appeared to be a relative maximum, as by that time already a substantial fraction of patients had passed away or had deteriorated in such a way that they did not want or were able to undergo another examination.

We tried to circumvent this by deriving diagnoses from the medical charts of those patients who were not diagnosed by our pair of specialists. Although our validation experiment showed that there was good agreement between these two methods of obtaining final diagnoses, we cannot exclude that it may have biased our results. Simultaneous SPECT scans,

which were reasonably accurate in diagnosing IPD in our study population, appear to confirm this relative lack of bias. In our population the FP-CIT SPECT scan did not reach a specificity of 100%, confirming an earlier report that a substantial fraction of early stage IPD patients have a normal SPECT scan [19].

We found substantially lower values for sensitivity and specificity of TCS to diagnose IPD than reported in earlier studies, including our own [20-49, [60]]. In diagnostic accuracy studies there are two major sources of variability: spectrum bias and test review bias [61]. Spectrum bias is the skewing of test parameters due to differences between study populations. Test review bias is skewing of test parameters due to differences in the amount of clinical information available to the investigator interpreting the test result. We think that spectrum bias is the main cause of the substantial differences between ours and earlier studies. With one exception [30], all the earlier studies were retrospective and involved patients who had already been diagnosed clinically with definite IPD. These later-stage patients are obviously not the patients for whom one needs additional diagnostic tools such as a TCS, as these patients already have a clinical diagnosis. Our study show that results obtained in already diagnosed patients cannot be simply extrapolated to early stage, as yet undiagnosed patients. One could even argue that more selective inclusion of those for whom the diagnosis would really be a 'toss-up', may provide different results (perhaps even lower sensitivity/specificity).

Our results also differ from the only other prospective study [30]. We believe here spectrum bias also plays a role: Gaenslen et al excluded patients with resting tremor, which we did not. The establishment of a definite diagnosis also differed between our studies. Gaenslen et al were not able to reach a definite diagnosis in all patients, possibly due to the shorter follow-up (1 vs. 2 years in our study).

We cannot rule out test review bias, as we did try to blind the TCS examiner, but not to great lengths. But this, if present, would have skewed the results of Gaenslen et al, and not ours, as we found less diagnostic accuracy in the TCS.

Both our examiners had more than twenty years of experience in ultrasound, and one of us (WHM) spent considerable time, for this research project, training with Prof. Berg's group in Tübingen, Germany. We had already done an inter-observer study, which yielded reasonable intra- and inter-rater reliability, in accordance with results by others [54, 62]. Results of TCS seem not be substantially influenced by the type of ultrasound device used [38], and we have in the past also found good diagnostic accuracy in later-stage IPD patients when studied retrospectively [60]. One might even reason that the fact that the investigators were so well-trained may imply that real-world utility would be even lower than found.

We thus feel that the crucial difference between earlier studies and ours is the prospective unselected nature of our patient population. Ours represented exactly the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. We show here that, in our hands, the TCS cannot be used reliably for that purpose.

Acknowledgements We would like to thank professor Daniela Berg, consultant neurologists Bert Anten, Fred Vreeling, Ania Winogrodzka and Marinus van Kroonenburgh for their helpful cooperation.

Authors' contributions AV, WW, WM, and AK had the idea and designed the protocol. Trial was done by AB, AV, WW, WM. Interpretation of the data was done by AB, WW, and AK. Paper was written by all authors.

Data sharing Dataset available from the corresponding author at Dryad repository, who will provide a permanent, citable and open access home for the dataset.

Conflict of Interest We have no conflicts of interest to declare.

Funding source This work was funded by the "Stichting Internationaal Parkinson Fonds, The Netherlands".

Legends to figures

Figure 1. Patiens flow chart

Figure 2. Boxplot of final diagnoses compared to the range of the maximum size of the hyperechogenic substantia nigra (SN+)

Figure 3. ROC Curve the maximum and the sum of the area of hyperechogenic substantia nigra (SN+) correlated with the final diagnosis Idiopathic Parkinson's Disease (IPD)

References

1. Kalra, S., D.G. Grosset, and H.T. Benamer, *Differentiating vascular parkinsonism from idiopathic Parkinson's disease: a systematic review.* Mov Disord, 2010. **25**(2): p. 149-56.
2. Diaz-Corrales, F.J., et al., *Clinical features and 123I-FP-CIT SPECT imaging in drug-induced parkinsonism and Parkinson's disease.* Eur J Nucl Med Mol Imaging, 2010. **37**(3): p. 556-64.
3. Wenning, G.K., et al., *What clinical features are most useful to distinguish definite multiple system atrophy from Parkinson's disease?* J Neurol Neurosurg Psychiatry, 2000. **68**(4): p. 434-40.
4. Osaki, Y., et al., *Accuracy of clinical diagnosis of progressive supranuclear palsy.* Mov Disord, 2004. **19**(2): p. 181-9.
5. Litvan, I., et al., *Accuracy of the clinical diagnosis of corticobasal degeneration: a clinicopathologic study.* Neurology, 1997. **48**(1): p. 119-25.
6. Wadia, P.M. and A.E. Lang, *The many faces of corticobasal degeneration.* Parkinsonism Relat Disord, 2007. **13 Suppl 3**: p. S336-40.
7. Litvan, I., et al., *Accuracy of the clinical diagnoses of Lewy body disease, Parkinson disease, and dementia with Lewy bodies: a clinicopathologic study.* Arch Neurol, 1998. **55**(7): p. 969-78.
8. Cardoso, F., S.T. Camargos, and G.A. Silva Junior, *Etiology of parkinsonism in a Brazilian movement disorders clinic.* Arq Neuropsiquiatr, 1998. **56**(2): p. 171-5.
9. Hughes, A.J., et al., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.* J Neurol Neurosurg Psychiatry, 1992. **55**(3): p. 181-4.
10. Munhoz, R.P., L.C. Werneck, and H.A. Teive, *The differential diagnoses of parkinsonism: findings from a cohort of 1528 patients and a 10 years comparison in tertiary movement disorders clinics.* Clin Neurol Neurosurg, 2010. **112**(5): p. 431-5.
11. Newman, E.J., et al., *Accuracy of Parkinson's disease diagnosis in 610 general practice patients in the West of Scotland.* Mov Disord, 2009. **24**(16): p. 2379-85.
12. Rajput, D.R., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease.* J Neurol Neurosurg Psychiatry, 1993. **56**(8): p. 938-9.
13. Ahlskog, J.E., *Diagnosis and differential diagnosis of Parkinson's disease and parkinsonism.* Parkinsonism Relat Disord, 2000. **7**(1): p. 63-70.
14. Caslake, R., et al., *Changes in diagnosis with follow-up in an incident cohort of patients with parkinsonism.* J Neurol Neurosurg Psychiatry, 2008. **79**(11): p. 1202-7.
15. Marshall, V.L., et al., *Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: a 3-year European multicenter study with repeat [123I]FP-CIT SPECT.* Mov Disord, 2009. **24**(4): p. 500-8.
16. McKinnon, J.H., et al., *Sniffing out Parkinson disease: can olfactory testing differentiate parkinsonian disorders?* Neurologist, 2007. **13**(6): p. 382-5.
17. Katzenschlager, R. and A.J. Lees, *Olfaction and Parkinson's syndromes: its role in differential diagnosis.* Curr Opin Neurol, 2004. **17**(4): p. 417-23.
18. Brooks, D.J., *Imaging approaches to Parkinson disease.* J Nucl Med, 2010. **51**(4): p. 596-609.
19. Vlaar, A.M., et al., *Meta-analysis of the literature on diagnostic accuracy of SPECT in parkinsonian syndromes.* BMC Neurol, 2007. **7**: p. 27.
20. Piccini, P. and A. Whone, *Functional brain imaging in the differential diagnosis of Parkinson's disease.* Lancet Neurol, 2004. **3**(5): p. 284-90.

21. Walter, U., et al., *Substantia nigra echogenicity is normal in non-extrapyramidal cerebral disorders but increased in Parkinson's disease.* J Neural Transm, 2002. **109**(2): p. 191-6.
22. Berg, D., C. Siefker, and G. Becker, *Echogenicity of the substantia nigra in Parkinson's disease and its relation to clinical findings.* J Neurol, 2001. **248**(8): p. 684-9.
23. Ressner, P., et al., *Hyperechogenicity of the substantia nigra in Parkinson's disease.* J Neuroimaging, 2007. **17**(2): p. 164-7.
24. Huang, Y.W., et al., *Transcranial imaging of substantia nigra hyperechogenicity in a Taiwanese cohort of Parkinson's disease.* Mov Disord, 2007. **22**(4): p. 550-5.
25. Spiegel, J., et al., *Transcranial sonography and [123I]FP-CIT SPECT disclose complementary aspects of Parkinson's disease.* Brain, 2006. **129**(Pt 5): p. 1188-93.
26. Becker, G., et al., *Degeneration of substantia nigra in chronic Parkinson's disease visualized by transcranial color-coded real-time sonography.* Neurology, 1995. **45**(1): p. 182-4.
27. Berg, D., et al., *Vulnerability of the nigrostriatal system as detected by transcranial ultrasound.* Neurology, 1999. **53**(5): p. 1026-31.
28. Berg, D., et al., *Echogenicity of the substantia nigra: association with increased iron content and marker for susceptibility to nigrostriatal injury.* Arch Neurol, 2002. **59**(6): p. 999-1005.
29. Kim, J.Y., et al., *Midbrain transcranial sonography in Korean patients with Parkinson's disease.* Mov Disord, 2007. **22**(13): p. 1922-6.
30. Gaenslen, A., et al., *The specificity and sensitivity of transcranial ultrasound in the differential diagnosis of Parkinson's disease: a prospective blinded study.* Lancet Neurol, 2008. **7**(5): p. 417-24.
31. Berg, D., et al., *Five-year follow-up study of hyperechogenicity of the substantia nigra in Parkinson's disease.* Mov Disord, 2005. **20**(3): p. 383-5.
32. Walter, U., et al., *Transcranial brain sonography findings in clinical subgroups of idiopathic Parkinson's disease.* Mov Disord, 2007. **22**(1): p. 48-54.
33. Walter, U., D. Skoloudik, and D. Berg, *Transcranial sonography findings related to non-motor features of Parkinson's disease.* J Neurol Sci, 2010. **289**(1-2): p. 123-7.
34. Schweitzer, K.J., et al., *Substantia nigra hyperechogenicity as a marker of predisposition and slower progression in Parkinson's disease.* Mov Disord, 2006. **21**(1): p. 94-8.
35. Weise, D., et al., *Substantia nigra echogenicity: A structural correlate of functional impairment of the dopaminergic striatal projection in Parkinson's disease.* Mov Disord, 2009. **24**(11): p. 1669-75.
36. Prestel, J., et al., *Predictive value of transcranial sonography in the diagnosis of Parkinson's disease.* Mov Disord, 2006. **21**(10): p. 1763-5.
37. Hagenah, J.M., et al., *Substantia nigra hyperechogenicity correlates with clinical status and number of Parkin mutated alleles.* J Neurol, 2007. **254**(10): p. 1407-13.
38. Mehnert, S., et al., *Transcranial sonography for diagnosis of Parkinson's disease.* BMC Neurol, 2010. **10**: p. 9.
39. Walter, U., et al., *Brain parenchyma sonography discriminates Parkinson's disease and atypical parkinsonian syndromes.* Neurology, 2003. **60**(1): p. 74-7.
40. Walter, U., et al., *Transcranial brain sonography findings in discriminating between parkinsonism and idiopathic Parkinson disease.* Arch Neurol, 2007. **64**(11): p. 1635-40.

41. Behnke, S., et al., *Differentiation of Parkinson's disease and atypical parkinsonian syndromes by transcranial ultrasound.* J Neurol Neurosurg Psychiatry, 2005. **76**(3): p. 423-5.
42. Okawa, M., et al., *Transcranial sonography of the substantia nigra in Japanese patients with Parkinson's disease or atypical parkinsonism: clinical potential and limitations.* Intern Med, 2007. **46**(18): p. 1527-31.
43. Ebentheuer, J., et al., *Substantia nigra echogenicity in progressive supranuclear palsy.* Mov Disord, 2010. **25**(6): p. 765-8.
44. Lauckaite, K., et al., *Specificity of transcranial sonography in parkinson spectrum disorders in comparison to degenerative cognitive syndromes.* BMC Neurol, 2012. **12**: p. 12.
45. Tsai, C.F., et al., *Transcranial color-coded sonography helps differentiation between idiopathic Parkinson's disease and vascular parkinsonism.* J Neurol, 2007. **254**(4): p. 501-7.
46. Doepp, F., et al., *Brain parenchyma sonography and 123I-FP-CIT SPECT in Parkinson's disease and essential tremor.* Mov Disord, 2008. **23**(3): p. 405-10.
47. Budisic, M., et al., *Distinguishing Parkinson's disease and essential tremor with transcranial sonography.* Acta Neurol Scand, 2009. **119**(1): p. 17-21.
48. Stockner, H., et al., *Midbrain sonography in patients with essential tremor.* Mov Disord, 2007. **22**(3): p. 414-7.
49. Walter, U., et al., *Sonographic discrimination of dementia with Lewy bodies and Parkinson's disease with dementia.* J Neurol, 2006. **253**(4): p. 448-54.
50. Walter, U., et al., *Sonographic discrimination of corticobasal degeneration vs progressive supranuclear palsy.* Neurology, 2004. **63**(3): p. 504-9.
51. Bouwmans, A.E., et al., *Transcranial sonography for the discrimination of idiopathic Parkinson's disease from the atypical parkinsonian syndromes.* Int Rev Neurobiol, 2010. **90**: p. 121-46.
52. Vlaar, A.M., et al., *Protocol of a prospective study on the diagnostic value of transcranial duplex scanning of the substantia nigra in patients with parkinsonian symptoms.* BMC Neurol, 2007. **7**: p. 28.
53. Berg, D., J. Godau, and U. Walter, *Transcranial sonography in movement disorders.* Lancet Neurol, 2008. **7**(11): p. 1044-55.
54. Vlaar, A., et al., *The Reliability of Transcranial Duplex Scanning in Parkinsonian Patients: Comparison of Different Observers and Ultrasound Systems.* Ultraschall Med, 2010.
55. *The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations.* Mov Disord, 2003. **18**(7): p. 738-50.
56. Gilman, S., et al., *Second consensus statement on the diagnosis of multiple system atrophy.* Neurology, 2008. **71**(9): p. 670-6.
57. Litvan, I., et al., *Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop.* Neurology, 1996. **47**(1): p. 1-9.
58. McKeith, I.G., et al., *Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium.* Neurology, 2005. **65**(12): p. 1863-72.
59. Mahapatra, R.K., et al., *Corticobasal degeneration.* Lancet Neurol, 2004. **3**(12): p. 736-43.
60. Vlaar, A.M., et al., *The predictive value of transcranial duplex sonography for the clinical diagnosis in undiagnosed parkinsonian syndromes: comparison with SPECT scans.* BMC Neurol, 2008. **8**: p. 42.

61. Bachmann, L.M., et al., *Multivariable adjustments counteract spectrum and test review bias in accuracy studies.* J Clin Epidemiol, 2009. **62**(4): p. 357-361 e2.
62. van de Loo, S., et al., *Reproducibility and diagnostic accuracy of substantia nigra sonography for the diagnosis of Parkinson's disease.* J Neurol Neurosurg Psychiatry, 2010. **81**(10): p. 1087-92.

Specificity and sensitivity of transcranial sonography of the substantia nigra in the diagnosis of Parkinson's disease: prospective cohort study in 196 patients

Angela E. P. Bouwmans MD [a], Annemarie M. M. Vlaar MD PhD[b], Werner H. Mess Prof MD PhD[c], Alfons Kessels MD MSc[d], and Wim E. J. Weber MD PhD [a]

Departments of [a]Neurology, of [c]Clinical Neurophysiology, and of [d]Clinical Epidemiology and Medical Technology Assesment (KEMTA), Maastricht University Medical Centre, Maastricht, The Netherlands

[b]Department of Neurology, Sint Lucas Andreas Hospital, Amsterdam, The Netherlands

Corresponding author: dr. W.E.J. Weber, Dept. of Neurology, Maastricht University Medical Centre, PO Box 5800, 6202 AZ Maastricht, The Netherlands

Tel.: +31-43-/3875058

Fax. : +31-43-/3877055

E-mail : wim.weber@mumc.nl

Keywords: transcranial sonography, substantia nigra, Parkinson's disease, diagnostic accuracy

Word count: 3989

Abstract Numerous ultrasound studies have suggested that a typical enlarged area of echogenicity in the substantia nigra (SN+) can help diagnose idiopathic Parkinson's disease (IPD). Almost~~However almost~~ all these studies were retrospective and involved patients with well-established diagnoses and long disease duration.

Objective Assessment of the diagnostic accuracy of transcranial sonography (TCS) of the substantia nigra in the patient with an undiagnosed parkinsonian syndrome of recent onset.

Design Prospective cohort study for diagnostic accuracy.

Setting Neurology outpatient clinics of two teaching hospitals in the Netherlands.

Patients 196 consecutive patients, who were referred to two neurology outpatient clinics for analysis of clinically unclear parkinsonism. Within two weeks of inclusion all patients also underwent a TCS and a ^{123}I-ioflupane Single Photon Emission Computer Tomography (FP-CIT SPECT) scan of the brain (n=176).

Outcome measures After two years, patients were re-examined by two movement disorder specialist neurologists for a final clinical diagnosis, that served as a surrogate gold standard for our study.

Results Temporal acoustic windows were insufficient in 45 of 241 patients (18.67%). The final clinical diagnosis was IPD in 102 (52,0%) patients. Twenty-four (12,3%) patients were diagnosed with atypical parkinsonisms (APS) of which 8 (4,0%) multisystem atrophy (MSA), 6 (3,1%) progressive supranuclear palsy (PSP), 6 (3,1%) Lewy body dementia

(LBD) and 4 (2,-0%) corticobasal degeneration (CBD). Twenty-one (10,-7%) patients had a diagnosis of vascular parkinsonism (VP), 20 (10,-2%) essential tremor (ET), 7 (3,-6%) drug-induced parkinsonism (DIP) and 22 (11,-2%) patients had no parkinsonism but an ~~a~~ alternative diagnosis. The sensitivity of a SN+ for the diagnosis IPD was 0,-40 (Confidence Interval (CI) 0,-30-0,50) and the specificity 0,-61 (CI 0,-52-0,-70). Hereby the positive predictive value (PPV) was 0,-53 and the negative predictive value (NPV) 0,-48. The sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was 0.88 (CI 0,-~~8~~1-0,-95) and 0.68 (CI 0,-58-0,-76) with a PPV of 0,-75 and a NPV of 0,-84.

Conclusion The diagnostic accuracy of TCS in early stage Parkinson's disease is not sufficient for routine clinical use.

Clinicaltrials.gov identifier: NCT0036819

Article summary

Article focus

- We wanted to assess the diagnostic accuracy of transcranial sonography (TCS) of the substantia nigra in patients with an undiagnosed parkinsonian syndrome of recent onset.
- A large body of evidence suggests that TCS of the substantia nigra can help diagnose idiopathic Parkinson's disease (IPD). The problem is that almost all these studies were retrospective and involved patients with well-established diagnoses and long disease duration.

Key messages

- The diagnostic accuracy of TCS in early stage Parkinson's disease is not sufficient for routine clinical use.

Strengths and limitations of this study

Strength of our study is its guaranteed prospective nature: we registered this study prospectively and we carried it out exactly as proposed in the published protocol. It is the largest prospective study on this technique in this patient population up till now. At inclusion we excluded the patients with already a clear diagnosis, thus closely mimicking the clinical situation in which the neurologist would need an additional tool for diagnostic workup.

A limitation, as in all these studies, is the lack of an objective gold standard, i.e. neuropathological analysis. We used clinical diagnosis after 2 years follow-up as gold standard. Longer follow-up periods will probably increase diagnostic accuracy, but will also lead to higher attrition rates in these elderly populations.

Introduction

In clinical practice the diagnosis of idiopathic Parkinson's disease (IPD), delineating it from the atypical parkinsonisms (APS), vascular parkinsonism (VP), drug induced parkinsonism (DIP), and essential tremor (ET) is still difficult[1-8]. Especially in the early stage of these diseases a large group of patients is erroneously diagnosed, even by experienced movement disorder specialists, when one uses post-mortem findings as a gold standard[9-13]. Longer-term follow-up studies with clinical criteria as a gold standard found that IPD was frequently overdiagnosed initially[14, 15]. As these disorders have~~demand vastly differing therapies along~~ varying prognoses, a multitude of ancillary investigations has been proposed as aids in the early diagnosis of IPD[16-20]. Of all these, ^{123}I-ioflupane Single Photon Emission Computer Tomography (FP-CIT SPECT) scans are most widely used in routine clinical practice to diagnose IPD. But a substantial fraction of patients with early IPD have normal scans, and the costs and use of intravenous radio-active tracers are seen as important disadvantages of this technique[19].

The search for a cheaper and more patient-friendly technique to diagnose IPD has thus continued and over the last 10 years transcranial sonography (TCS) of the substantia nigra (SN) has emerged as a promising tool in this regard. Numerous ultrasound studies have found that a significant percentage of patients with IPD has a typical enlarged area of echogenicity in the substantia nigra (SN+), which is thought to be associated with increased iron concentrations [21-38]. Some of these studies have suggested that with this echofeature one can diagnose IPD with reasonable sensitivity and specificity. Further research along these lines found that TCS might also be used to delineate IPD from the APS [39-44], such as Multiple System Atrophy (MSA) and Progressive Supranuclear Palsy (PSP). These patients appear to have normal or only a moderately enhanced hyperechogenic SN as have patients with VP [45], ET [46-48] and DIP. Patients with Lewy Body dementia (LBD) [49] and

Cortical Basal Degeneration (CBD) [50] have been reported to share the same echofeature with IPD patients, and researchers have found that the accuracy of the differential diagnosis can be enhanced by additional assessments of the echogenicity of the basal ganglia. Hyperechogenicity of the lentiform nucleus is commonly seen in patients with CBD, whereas patients with IPD have this echofeature only rarely. Furthermore, research showed that the absence of bilateral marked SN+ discriminated IPD from LBD with a moderate to good sensitivity, and a good specificity and positive predictive value [49]. All these different findings combined could then give a ´diagnostic fingerprint´ for these disorders by following an algorithm we recently postulated [51].

However almost all studies were retrospective and involved patients with well-established diagnoses and long disease duration. These findings can thus not simply be extrapolated to the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. Up till now only one prospective study has assessed the diagnostic accuracy in patients with recent onset parkinsonian signs and symptoms [30]. This study was relatively small, excluded patients with tremor, and followed up patients for only 12 months.

~~We have now assessed the diagnostic accuracy of TCS of the SN in 196 patients referred by their general practitioner (GP) for analysis of a parkinsonian syndrome of recent onset. We used a clinical diagnosis after two years as a surrogate gold standard and also compared TCS with FP-CIT-SPECT scans.~~

Methods

Patients

This was a prospective study testing the diagnostic accuracy of TCS of the SN in patients who are referred by their GP for a first consultation by a neurologist because of recent-onset parkinsonism of unclear origin [52]. The Institutional Review Board (IRB) of the University Hospital Maastricht approved the study (MEC 05–228, 4 April 2006), and the study was registered prospectively under (ITRSCC) NCT0036819. The study protocol was published before the study started ~~(weblink: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2034584)~~ [52].

We considered 283 consecutive patients, who were referred to two neurology outpatient clinics for analysis of clinically unclear parkinsonism (Neurology Outpatient Clinic of the Maastricht University Medical Centre (MUMC) in Maastricht and the Orbis Medical centre in Sittard, The Netherlands). Patients, in whom a definite diagnosis could be made at the first visit, were excluded from the study (n=42). Hence, we enrolled 241 patients. After signing informed consent, upon entering the study, all subjects underwent a structured interview and a neurological examination (See Additional file 1 [52]). These tests were performed by a physician not treating the patient and blinded for information in the routine clinical records [52].

Within two weeks of inclusion all patients underwent a TCS of the SN, at the department of Neurophysiology of the two mentioned hospitals. In each hospital TCS was done by one specially trained investigator (P. Wuisman MD in Orbis Medical Centre, Sittard, and Prof. W. Mess (WHM) in the MUMC). WHM is a very experienced sonographer, who did additional training with Prof. D. Berg one of the pioneers of this technique [53]. To ensure validity of the TCS assessments among our two sonographers we had already done an interobserver

study, and found an acceptable interobserver agreement with kappa values in the 0.7- 0.8 range [54].

Patients in whom a TCS of the SN was not possible because of a non-accessible bone window were excluded from the study resulting eventually in a group of 196 patients. Within two weeks of inclusion all patients also underwent a FP-CIT SPECT scan of the brain as described in our protocol [52].

After two years, patients were re-examined by two movement disorder specialist neurologists for a final clinical diagnosis, that served as a surrogate gold standard for our study. The four consultant neurologists who alternatingly did these assessments were all specialists in movement disorders with more than ten years' experience in this field (Bert Anten MD PhD, Fred Vreeling MD PhD, Wim Weber MD PhD, and Ania Winogrodzka MD PhD). These investigators were blinded for all test results of these patients. In the planning of these visits we had made sure that neither one of the two neurologists had ever seen the patient. They were asked to interview and examine the patient, as they would normally do during a routine neurologic consultation. They were asked to fill out the same standard form as had been done by the including investigator during the first visit of the patient (see Additional file 1). Among other items this form contained the Unified Parkinson's Disease Rating Scale (UPDRS)-III score [55], and afterwards the neurologists received these scores of the patient at the first visit, so that they could evaluate whether the patient had had any progression on that scale. They also received the results of the brain scan, preferable a Magnetic Resonance Imaging (MRI) of the cerebrum however when not possible due to claustrophobia or devices not allowed in the MRI, a Computer Tomography (CT) of the brain. Each neurologist was asked then to reach a final clinical diagnosis of the parkinsonian syndrome using the diagnostic clinical criteria for IPD and APS [9, 56-59]. One investigator compared these scores and when there was no agreement, the two neurologists were asked to discuss these patients using their notes, in an

effort to reach agreement on the final diagnosis. In all cases except five patients, this discussion resulted in agreement on the final neurological diagnosis. Concerning the five patients to discussion, the diagnosis made at regular controls on the outpatient clinic of Neurology was taken as a third opinion and so a final diagnosis had been made.

TCS

One investigator per hospital and blinded to clinical information, did the ultrasound imaging (sonography) with~~ultrasounds using~~ a SONOS 5500 (Philips, Eindhoven, The Netherlands). The examination took place in a darkened room with the patient already lying on the examination table before the investigator entered the room. This was done~~in order~~ to minimize the possible identification of a patient's clinical signs. Patient and investigator had been asked not to talk about medical information.

TCS investigation was performed bilaterally through the pre-auricular bone window with a 2–4 MHz phased array transducer. The quality of the bone window was scored as good, moderate or inferior. Two different methods were applied for the evaluation of the SN. First, the presence or absence of an obviously visible SN was scored (qualitative method). Second, the area of an possible signal intensity was manually encircled and automatically calculated (quantitative method). This was only done when the increase of the hyperechogenicity was located in the anatomical distribution of the SN meaning showing a typically stripe-shaped configuration. Both the right and left SN were measured from both sides.

FP CIT SPECT

The SPECT scanning was done ~~had been performed~~ within 2 weeks of inclusion in the study. In this study FP-CIT (^{123}I-ioflupane, Nycomed, Amersham, U.K.) was~~is~~ used as presynaptic radiotracer. Medication (amphetamine, citalopram, fentanyl, fluoxetine, fluvoxamine,

paroxetine, sertraline, venlafaxine) which could interfere with the radiotracer had been discontinued at least 5 half life times. After intravenous injection of the tracer, SPECT measured baseline dopamine transporter integrity in the brain. SPECT was performed with a triple head camera (MultiSPECT3, Siemens, Ohio, USA) equipped with high-resolution collimators. A semi-automatic template model program was used to calculate the ratios between left striatal and right striatal and occipital regions respectively. Total time of acquisition was 30 minutes (45 seconds per frame for 40 views per detector). Zoom factor: 1.00 and the matrix size: 128 × 128. Filtered back-projection acquisition was performed. Images were filtered using a Butterworth filter with a cut-off value of: 0.4–0.5 and an order of 5. A division between the caudate nucleus and putamen was made. The ratios were corrected using Alderson's brain phantom, with known activities in the caudate nucleus and putamen. A binding of two standard deviations below healthy controls was considered as abnormal (FP-CIT 8.25 +-1.85 for putamen and 7.76 +-1.77 for caudate nucleus). Beside quantitative the scans were also judged visually by the same nuclear specialist blinded for the final clinical diagnosis. If quantitative and visual judgments did not agree the conclusion of visual judgment was taken.

Statistical analysis

Analyses were performed with SPSS, version 16.0. To determine the diagnostic performance of the SN+ and the FP-CIT SPECT we constructed Receiver Operating Characteristics (ROC) curves and calculated the Area under The Curve (AUC) and their p-values.

Role of the funding source

None.

Results

Patient characteristics

We had originally included 241 patients into the TCS study after approaching 283 possible candidates (See patients flowchart) in the period September 2006 until September 2008. The number of patients with no accessible temporal bone window was 45 (18,7%); these were slightly older (mean age of 72,4 versus 69,2 years) and there were by far more females (71% versus 26%) in this group compared to group of patients included in the study . This resulted in a group of 196 patients who had undergone an initial TCS. After two years 30 (15,3%) patients had died and 52 (26,5%) patients were not able or willing to undergo a second neurologic examination. The remaining 114 patients all underwent examination by two neurologists for a final clinical diagnosis of their movement disorder in the period September 2008 until September 2010. For the other 82 patients we derived a clinical diagnosis from the most recent clinical charts by the treating neurologist. To check the validity of this approach, we also derived these diagnoses from medical records for the 114 patients of whom we did have a gold standard diagnosis, and we found an agreement between these diagnoses with a kappa of 0.8. We also found no significant differences in the distribution of diagnoses between the patients groups with and without a gold standard diagnosis. But the group without the gold standard follow-up diagnosis did have a significantly higher age (70,5 versus 67,7 years, p= 0,034) and a higher UPDRS total score at inclusion (30,2 versus 22,8, p= 0,031).

The final clinical diagnosis was IPD in 102 (52,0%) patients. See for further division of the division of the final diagnoses table 1. The remaining 22 (11,2%) patients with no parkinsonism had alternative diagnoses like isolated tremor, orthostatic tremor, tardive dyskinesia, multi-infarction dementia, Alzheimer's disease~~M. Alzheimer~~, stroke, hypoxic encephalopathy, and psychogenic disorders.

Table 1. Patient characteristics

	All patients (n=196)	IPD (n=102)	APS (n=24)	VP (n=21)	ET (n=20)	DIP (n=7)	No parkinsonism (n=22)
Mean age in years (SD)	69,-2 (9,-54)	68,-5 (9,-3)	69,-6 (8,-6)	76,-3ζ (5,-9)	69,-4 (11,-2)	63,-1 (10,-4)	67,-2 (10,-1)
Men in %	74,-0%	71,-6%	79,-2%	85,-7%	75,-0%	85,-7%	63,-6%
Mean duration complaints in months (SD)	34,-2 (43,-57)	29,-8 (41,-7)	25,-8 (20,-8)	25,-0 (22,-4)	66,-2† (56,-1)	68,-6 (61,-1)	32,-3 (52,-6)
Mean score UPDRS-III at inclusion (SD)	13,-7 (7,-3)	13,-2 (6,-1)	17,-8 (9,-6)	17,-7ζζ (8,-6)	10,-7†† (5,-3)	14,-9 (5,-1)	9,-9 (6,-8)

IPD = Parkinson's disease, APS = atypical parkinsonian syndromes, VP = vascular parkinsonism, ET = essential tremor, DIP = drug induced parkinsonism, UPDRS-III = Unified Parkinson's Disease Rating Scale part III, ζ = significant (p value below 0.05) higher age compared to ET, DIP and no parkinsonism, ζζ = significant (p value below 0.05) higher UPDRS-III compared to IPD, †† = significant (p value below 0.05) lower UPDRS-III compared to APS and VP, † = significant (p value below 0.05) longer mean duration of complaints compared to IPD, APS and VP

Final diagnoses and SN

Table 2 gives the presence or absence of a SN+ related to the final diagnoses. The cut-off of 0.20 cm2 corresponds to the 75th percentile of hyperechogenic signal extent at the SN in a healthy population [21, 27, 39].

Table 2. Final diagnoses divided to the results of the transcranial sonography (TCS)

SN	IPD no.	MSA no.	PSP no.	LBD no.	CBD no.	VP no.	ET no.	DIP no.	No Parkinsonism no.	Total no.
~~No presence of hyperechogenic~~ SN- (%) ~~0,20 cm2 or more~~	61 (60)	4 (50)	2 (33)	3 (50)	2 (50)	12 (57)	13 (65)	5 (71)	16 (73)	118 (60)
~~Presence of hyperechogenic~~ SN+ (%) ~~0,20 cm2 or more~~	41 (40)	4 (50)	4 (67)	3 (50)	2 (50)	9 (43)	7 (35)	2 (29)	6 (27)	78 (40)
Total	102	8	6	6	4	21	20	7	22	196

SN= substantia nigra, SN- = no presence of hyperechogenic SN 0,20 cm2 or more, SN+ = presence of hyperechogenic SN 0,20 cm2 or more , IPD = Parkinson's disease, MSA = multisystem atrophy, PSP = progressive supranuclear palsy, LBD = Lewy body dementia, VP = vascular parkinsonism, ET = essential tremor, DIP = drug induced parkinsonism

One can see that the presence and absence of the hyperechogenic SNs are distributed at random over the various diagnoses, without any preference for one particular diagnosis. We also found no significant difference for the maximum size or the sum of the area of the SN+ in the different diagnoses (see table 3, figure 1 and 2).

Table 3. Final diagnoses divided to the results of the FP-CIT SPECT

	IPD no.	MSA no.	PSP no.	LBD no.	CBD no.	VP no.	ET no.	DIP no.	No Parkinsonism no.	Total no.
Normal	11	3	2	1	1	11	18	6	16	69
Abnormal	80	4	4	4	3	7	1	1	3	107
Total	91	7	6	5	4	18	19	7	19	176

IPD= Parkinson's disease, MSA= multisystem atrophy, PSP= progressive supranuclear palsy, LBD= Lewy body dementia, VP= vascular parkinsonism, ET= essential tremor, DIP= drug induced parkinsonism

The maximum size of the area of the SN+ is the one side of the mesencephalon on which the SN+ is the largest one. In a considerable numberamount of patients the SN+ was is bilaterally present, and then so the both areas of the SN+ were added. We were not able to obtain better diagnostic discrimination with other TCS cut-offs. E.g., when we lowered the sensitivity threshold to an absolute minimum of 0.7, we obtained a cut-off of 0,3 cm2, but then specificity was 0.29.

are summed up resulting in the sum of the area. The mean area of the ROC curve wasis 0,541. The sensitivity of a SN+ for the diagnosis IPD was 0,40 (Confidence Interval (CI) 0,30-0,50) and the specificity 0,61 (CI 0,52-0,70). PositiveHereby the positive predictive value (PPV) was 0,53 and the negative predictive value (NPV) 0,48. Because of earlier research suggested that SN+ can help diagnose literature suggesting also a SN+ in the diagnoses LBD and CBS, we added these two groups to the IPD and recalculated, reachingcame to the same sensitivity (0,41) and specificity (0,62).

Earlier research had suggested that the symmetry of the SN+ helps to differentiate between IPD and LBD, so we also analysed this feature. Of the 3 SN+ in the patients with LBD, 2 (67%) were bilaterally hyperintense. However, 29 (71%) of the 41 SN+ in the IPD patients, were bilaterally hyperintense, so in our population this echofeature had no diagnostic

discriminatory value between the diagnoses LBD and IPD. As the current view is that these diagnoses clinically overlap, it may be inappropriate to consider the diagnosis of DLB instead of PD to be a diagnostic error.

Final diagnoses and FP-CIT-SPECT scan results

176 patients also underwent a FPCIT-SPECT at initial work-up, around the same time when they underwent a TCD (Table~~, see table~~ 3). The sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was respectively 0.88 (CI 0,81-0,95) and 0.68 (0,58-0,76) with a PPV of 0.75 and a NPV of 0.84. Figure 3 shows the ROC Curve of FP-CIT-SPECT minimal uptake in the putamen and nucleus caudatus as a diagnostic performance to detect IPD. ~~This was much better concerning the SPECT with a mean area of the ROC curve of 0,815 compared to the TCS.~~ TCS findings were concordant with SPECT findings in 89 of 176 patients (p= 0,36).

Table 3. Final diagnoses divided to the results of the FP CIT SPECT

FP CIT SPECT	IPD no.	MSA no.	PSP no.	LBD no.	CBD no.	VP no.	ET no.	DIP no.	No Parkinsonism no.	Total no.
Normal (%)	11 (12)	3 (43)	2 (33)	1 (20)	1 (25)	11 (61)	18 (95)	6 (86)	16 (84)	69 (39)
Abnormal (%)	80 (88)	4 (57)	4 (67)	4 (80)	3 (75)	7 (39)	1 (5)	1 (14)	3 (16)	107 (61)
Total	91	7	6	5	4	18	19	7	19	176

IPD= Parkinson's disease, MSA= multisystem atrophy, PSP= progressive supranuclear palsy, LBD= Lewy body dementia, VP= vascular parkinsonism, ET= essential tremor, DIP= drug induced parkinsonismIn the 114 patients which had been re-examined after a follow-up of two years, the SPECT scan and the TCS results were in agreement in only 50 patients (p= 0.53).

This concordance of TCS and SPECT data was randomly distributed over the diagnostic groups. We also studied diagnostic accuracy in terms of delineating PD from non-parkinsonian (APS) syndromes. When we grouped all IPD diagnoses together with all APS diagnoses vs. the rest, i.e. ET, DIP, VP, etc., we found similar specificity and sensitivity of TCS, respectively 0.67 and 0.43. In this analysis sensitivity of SPECT remained 0.84 and specificity increased to 0.84.

Discussion

We have tried to assess the diagnostic accuracy of TCS in IPD, in the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian syndrome that cannot be diagnosed clinically at the first visit. We thus assessed its accuracy in 241 consecutive patients referred by their GP for analysis of a parkinsonian syndrome of recent onset. We used a clinical diagnosis after two years as a surrogate gold standard and

also compared TCS with FP-CIT-SPECT scans. ~~We found no significant correlation between the SN+ and any of the final diagnoses in patients presenting with first symptoms of a hitherto undiagnosed parkinsonism.~~ Sensitivity and specificity of SN+ for the diagnosis of IPD was 0.4 and 0.61 respectively. Positive~~Hereby the positive~~ predictive value (PPV) was 0.53 and the negative predictive value (NPV) 0.48. In contrast, we found that the sensitivity and specificity of FP-CIT SPECT scans for diagnosing IPD was respectively 0.88 and 0.68 with a PPV of 0.75 and a NPV of 0.84. In our hands, temporal acoustic windows were insufficient in a relatively high proportion of patients: 18.67%.

Strength of our study is its guaranteed prospective nature: we registered this study prospectively and we carried it out exactly as proposed in the published protocol [52]. Another strength is its size: it is the largest prospective study on this technique in this patient population up till now. At inclusion we excluded the patients with already a clear diagnosis ~~which has not been done in the study of Gaenslen~~. We have also tried to obtain the best possible surrogate gold standard clinical diagnosis. We did this by having our patients examined by a pair of independent experienced movement disorder specialists. The accepted gold standard is postmortem neuropathological examination, but this is hardly feasible anymore in modern times, as relatives are reluctant to give permission for this. So, the methodologically highest achievable gold standard is clinical examination after several years. This follow-up is essential as the diagnostic criteria contain several items that can ~~instead of using~~ only be assessed after a certain amount of time (levodopa response, progression, other diagnoses. The ~~the results of the imaging techniques done by Gaenslen. We also tried to increase its validity by observing a~~ follow-up of two years. ~~This~~ appeared to be a relative maximum, as by that time already a substantial fraction of patients had passed away or had deteriorated in such a way that they did not want or were able to undergo another examination.

~~However, the size of our study population is about three times as large compared to the group of Gaenslen. Apart from the real gold standard of the post-mortem examination, which seems less feasible in modern times, we think that this gold standard diagnosis of IPD is methodologically the highest achievable one. Implicitly, the follow-up is also a weakness, as it led to considerable attrition.~~ We tried to circumvent this by deriving diagnoses from the medical charts of those patients who were not diagnosed by our pair of specialists. Although our validation experiment showed that there was good agreement between these two methods of obtaining final diagnoses, we cannot exclude that it may have biased our results. Simultaneous SPECT scans, which were reasonably accurate in diagnosing IPD in our study population, appear to confirm this relative lack of bias. In our population the FP-CIT SPECT scan did not reach a specificity of 100%, confirming an earlier report that a substantial fraction of early stage IPD patients have a normal SPECT scan~~scans~~ [19].

We found substantially lower values for sensitivity and specificity of TCS to diagnose IPD than reported in earlier studies, including our own [20-49, [60]]. In diagnostic accuracy studies there are two major sources of variability: spectrum bias and test review bias [61]. Spectrum bias is the skewing of test parameters due to differences between study populations. Test review bias is skewing of test parameters due to differences in the amount of clinical information available to the investigator interpreting the test result. We think that spectrum bias is the main cause of the substantial differences between ours and earlier studies. With one exception [30], all the earlier studies were retrospective and involved patients who had already been diagnosed clinically with definite IPD. These later-stage patients are obviously not the patients for whom one needs additional diagnostic tools such as a TCS, as these patients already have a~~an unequivocal~~ clinical diagnosis. Our study ~~clearly~~ show that results obtained in already diagnosed patients cannot be simply extrapolated to early stage, as yet undiagnosed patients. One could even argue that more selective inclusion of those for whom

the diagnosis would really be a 'toss-up', may provide different results (perhaps even lower sensitivity/specificity).

Our results also differ from the only other prospective study [30]. We believe here spectrum bias also plays a role: Gaenslen et al excluded patients with resting tremor, which we did not. The establishment of a definite diagnosis also differed between our studies. Gaenslen et al were not able to reach a definite diagnosis in all patients, possibly due to the shorter follow-up (1 vs. 2 years in our study).

We cannot rule out test review bias, as we did try to blind the TCS examiner, but not to great lengths. But this, if present, would have skewed the results of Gaenslen et al, and not ours, as we found less diagnostic accuracy in the TCS.

~~One could then argue that our TCS examiners were not experienced enough.~~ Both our examiners had more than twenty years of experience in ultrasound, and one of us (WHM) spent considerable time, for this research project, training with Prof. Berg's group in Tübingen, Germany. We ~~. As stated above, we~~ had already done an inter-observer study, which yielded reasonable intra- and inter-rater reliability, in accordance with results by others [54, 62]. Results of TCS seem not be substantially influenced by the type of ultrasound device used [38], and we have in the past also found good diagnostic accuracy in later-stage IPD patients when studied retrospectively [60]. One might even reason that the fact that the investigators were so well-trained may imply that real-world utility would be even lower than found.

We thus feel that the crucial difference between earlier studies and ours is the prospective unselected nature of our patient population. Ours represented exactly the clinical situation for which one would need the TCS, namely the patient with a recent-onset parkinsonian

syndrome that cannot be diagnosed clinically at the first visit. We show here that, in our hands, the TCS cannot be used reliably for that purpose.

Acknowledgements We would like to thank professor Daniela Berg, consultant neurologists Bert Anten, Fred Vreeling, Ania Winogrodzka and Marinus van Kroonenburgh for their helpful cooperation.

Authors' contributions AV, WW, WM, and AK had the idea and designed the protocol. Trial was done by AB, AV, WW, WM. Interpretation of the data was done by AB, WW, and AK. Paper was written by all authors.

Data sharing Dataset available from the corresponding author at Dryad repository, who will provide a permanent, citable and open access home for the dataset.

Conflict of Interest We have no conflicts of interest to declare.

Funding source This work was funded by the "Stichting Internationaal Parkinson Fonds, The Netherlands".

Legends to figures

Figure 1. Patiens flow chart

Figure 2. Boxplot of final diagnoses compared to the range of the maximum size of the hyperechogenic substantia nigra (SN+)

Figure 3. ROC Curve the maximum and the sum of the area of hyperechogenic substantia nigra (SN+) correlated with the final diagnosis Idiopathic Parkinson's Disease (IPD)

References

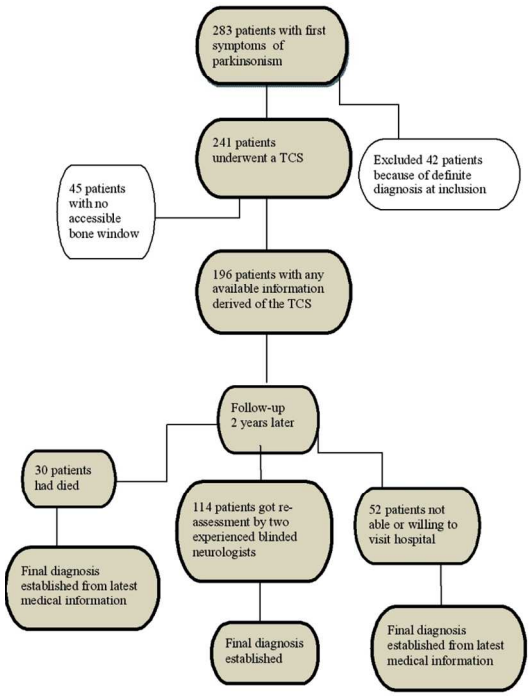
1. Kalra, S., D.G. Grosset, and H.T. Benamer, *Differentiating vascular parkinsonism from idiopathic Parkinson's disease: a systematic review.* Mov Disord, 2010. **25**(2): p. 149-56.
2. Diaz-Corrales, F.J., et al., *Clinical features and 123I-FP-CIT SPECT imaging in drug-induced parkinsonism and Parkinson's disease.* Eur J Nucl Med Mol Imaging, 2010. **37**(3): p. 556-64.
3. Wenning, G.K., et al., *What clinical features are most useful to distinguish definite multiple system atrophy from Parkinson's disease?* J Neurol Neurosurg Psychiatry, 2000. **68**(4): p. 434-40.
4. Osaki, Y., et al., *Accuracy of clinical diagnosis of progressive supranuclear palsy.* Mov Disord, 2004. **19**(2): p. 181-9.
5. Litvan, I., et al., *Accuracy of the clinical diagnosis of corticobasal degeneration: a clinicopathologic study.* Neurology, 1997. **48**(1): p. 119-25.
6. Wadia, P.M. and A.E. Lang, *The many faces of corticobasal degeneration.* Parkinsonism Relat Disord, 2007. **13 Suppl 3**: p. S336-40.
7. Litvan, I., et al., *Accuracy of the clinical diagnoses of Lewy body disease, Parkinson disease, and dementia with Lewy bodies: a clinicopathologic study.* Arch Neurol, 1998. **55**(7): p. 969-78.
8. Cardoso, F., S.T. Camargos, and G.A. Silva Junior, *Etiology of parkinsonism in a Brazilian movement disorders clinic.* Arq Neuropsiquiatr, 1998. **56**(2): p. 171-5.
9. Hughes, A.J., et al., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.* J Neurol Neurosurg Psychiatry, 1992. **55**(3): p. 181-4.
10. Munhoz, R.P., L.C. Werneck, and H.A. Teive, *The differential diagnoses of parkinsonism: findings from a cohort of 1528 patients and a 10 years comparison in tertiary movement disorders clinics.* Clin Neurol Neurosurg, 2010. **112**(5): p. 431-5.
11. Newman, E.J., et al., *Accuracy of Parkinson's disease diagnosis in 610 general practice patients in the West of Scotland.* Mov Disord, 2009. **24**(16): p. 2379-85.
12. Rajput, D.R., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease.* J Neurol Neurosurg Psychiatry, 1993. **56**(8): p. 938-9.
13. Ahlskog, J.E., *Diagnosis and differential diagnosis of Parkinson's disease and parkinsonism.* Parkinsonism Relat Disord, 2000. **7**(1): p. 63-70.
14. Caslake, R., et al., *Changes in diagnosis with follow-up in an incident cohort of patients with parkinsonism.* J Neurol Neurosurg Psychiatry, 2008. **79**(11): p. 1202-7.
15. Marshall, V.L., et al., *Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: a 3-year European multicenter study with repeat [123I]FP-CIT SPECT.* Mov Disord, 2009. **24**(4): p. 500-8.
16. McKinnon, J.H., et al., *Sniffing out Parkinson disease: can olfactory testing differentiate parkinsonian disorders?* Neurologist, 2007. **13**(6): p. 382-5.
17. Katzenschlager, R. and A.J. Lees, *Olfaction and Parkinson's syndromes: its role in differential diagnosis.* Curr Opin Neurol, 2004. **17**(4): p. 417-23.
18. Brooks, D.J., *Imaging approaches to Parkinson disease.* J Nucl Med, 2010. **51**(4): p. 596-609.
19. Vlaar, A.M., et al., *Meta-analysis of the literature on diagnostic accuracy of SPECT in parkinsonian syndromes.* BMC Neurol, 2007. **7**: p. 27.
20. Piccini, P. and A. Whone, *Functional brain imaging in the differential diagnosis of Parkinson's disease.* Lancet Neurol, 2004. **3**(5): p. 284-90.

21. Walter, U., et al., *Substantia nigra echogenicity is normal in non-extrapyramidal cerebral disorders but increased in Parkinson's disease.* J Neural Transm, 2002. **109**(2): p. 191-6.
22. Berg, D., C. Siefker, and G. Becker, *Echogenicity of the substantia nigra in Parkinson's disease and its relation to clinical findings.* J Neurol, 2001. **248**(8): p. 684-9.
23. Ressner, P., et al., *Hyperechogenicity of the substantia nigra in Parkinson's disease.* J Neuroimaging, 2007. **17**(2): p. 164-7.
24. Huang, Y.W., et al., *Transcranial imaging of substantia nigra hyperechogenicity in a Taiwanese cohort of Parkinson's disease.* Mov Disord, 2007. **22**(4): p. 550-5.
25. Spiegel, J., et al., *Transcranial sonography and [123I]FP-CIT SPECT disclose complementary aspects of Parkinson's disease.* Brain, 2006. **129**(Pt 5): p. 1188-93.
26. Becker, G., et al., *Degeneration of substantia nigra in chronic Parkinson's disease visualized by transcranial color-coded real-time sonography.* Neurology, 1995. **45**(1): p. 182-4.
27. Berg, D., et al., *Vulnerability of the nigrostriatal system as detected by transcranial ultrasound.* Neurology, 1999. **53**(5): p. 1026-31.
28. Berg, D., et al., *Echogenicity of the substantia nigra: association with increased iron content and marker for susceptibility to nigrostriatal injury.* Arch Neurol, 2002. **59**(6): p. 999-1005.
29. Kim, J.Y., et al., *Midbrain transcranial sonography in Korean patients with Parkinson's disease.* Mov Disord, 2007. **22**(13): p. 1922-6.
30. Gaenslen, A., et al., *The specificity and sensitivity of transcranial ultrasound in the differential diagnosis of Parkinson's disease: a prospective blinded study.* Lancet Neurol, 2008. **7**(5): p. 417-24.
31. Berg, D., et al., *Five-year follow-up study of hyperechogenicity of the substantia nigra in Parkinson's disease.* Mov Disord, 2005. **20**(3): p. 383-5.
32. Walter, U., et al., *Transcranial brain sonography findings in clinical subgroups of idiopathic Parkinson's disease.* Mov Disord, 2007. **22**(1): p. 48-54.
33. Walter, U., D. Skoloudik, and D. Berg, *Transcranial sonography findings related to non-motor features of Parkinson's disease.* J Neurol Sci, 2010. **289**(1-2): p. 123-7.
34. Schweitzer, K.J., et al., *Substantia nigra hyperechogenicity as a marker of predisposition and slower progression in Parkinson's disease.* Mov Disord, 2006. **21**(1): p. 94-8.
35. Weise, D., et al., *Substantia nigra echogenicity: A structural correlate of functional impairment of the dopaminergic striatal projection in Parkinson's disease.* Mov Disord, 2009. **24**(11): p. 1669-75.
36. Prestel, J., et al., *Predictive value of transcranial sonography in the diagnosis of Parkinson's disease.* Mov Disord, 2006. **21**(10): p. 1763-5.
37. Hagenah, J.M., et al., *Substantia nigra hyperechogenicity correlates with clinical status and number of Parkin mutated alleles.* J Neurol, 2007. **254**(10): p. 1407-13.
38. Mehnert, S., et al., *Transcranial sonography for diagnosis of Parkinson's disease.* BMC Neurol, 2010. **10**: p. 9.
39. Walter, U., et al., *Brain parenchyma sonography discriminates Parkinson's disease and atypical parkinsonian syndromes.* Neurology, 2003. **60**(1): p. 74-7.
40. Walter, U., et al., *Transcranial brain sonography findings in discriminating between parkinsonism and idiopathic Parkinson disease.* Arch Neurol, 2007. **64**(11): p. 1635-40.

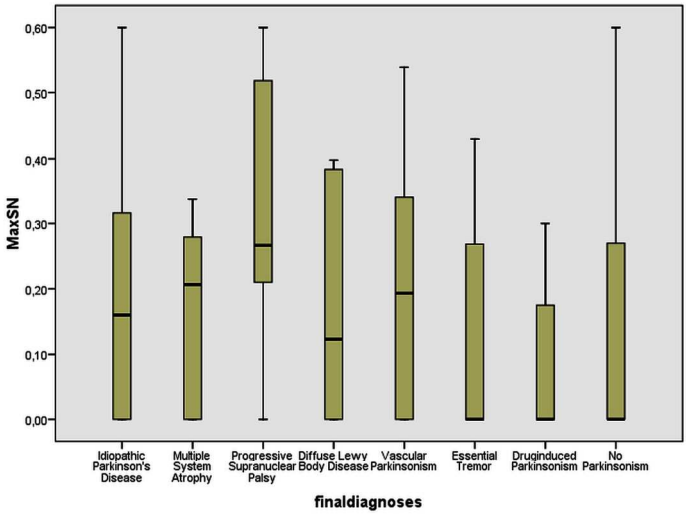
41. Behnke, S., et al., *Differentiation of Parkinson's disease and atypical parkinsonian syndromes by transcranial ultrasound.* J Neurol Neurosurg Psychiatry, 2005. **76**(3): p. 423-5.
42. Okawa, M., et al., *Transcranial sonography of the substantia nigra in Japanese patients with Parkinson's disease or atypical parkinsonism: clinical potential and limitations.* Intern Med, 2007. **46**(18): p. 1527-31.
43. Ebentheuer, J., et al., *Substantia nigra echogenicity in progressive supranuclear palsy.* Mov Disord, 2010. **25**(6): p. 765-8.
44. Lauckaite, K., et al., *Specificity of transcranial sonography in parkinson spectrum disorders in comparison to degenerative cognitive syndromes.* BMC Neurol, 2012. **12**: p. 12.
45. Tsai, C.F., et al., *Transcranial color-coded sonography helps differentiation between idiopathic Parkinson's disease and vascular parkinsonism.* J Neurol, 2007. **254**(4): p. 501-7.
46. Doepp, F., et al., *Brain parenchyma sonography and 123I-FP-CIT SPECT in Parkinson's disease and essential tremor.* Mov Disord, 2008. **23**(3): p. 405-10.
47. Budisic, M., et al., *Distinguishing Parkinson's disease and essential tremor with transcranial sonography.* Acta Neurol Scand, 2009. **119**(1): p. 17-21.
48. Stockner, H., et al., *Midbrain sonography in patients with essential tremor.* Mov Disord, 2007. **22**(3): p. 414-7.
49. Walter, U., et al., *Sonographic discrimination of dementia with Lewy bodies and Parkinson's disease with dementia.* J Neurol, 2006. **253**(4): p. 448-54.
50. Walter, U., et al., *Sonographic discrimination of corticobasal degeneration vs progressive supranuclear palsy.* Neurology, 2004. **63**(3): p. 504-9.
51. Bouwmans, A.E., et al., *Transcranial sonography for the discrimination of idiopathic Parkinson's disease from the atypical parkinsonian syndromes.* Int Rev Neurobiol, 2010. **90**: p. 121-46.
52. Vlaar, A.M., et al., *Protocol of a prospective study on the diagnostic value of transcranial duplex scanning of the substantia nigra in patients with parkinsonian symptoms.* BMC Neurol, 2007. **7**: p. 28.
53. Berg, D., J. Godau, and U. Walter, *Transcranial sonography in movement disorders.* Lancet Neurol, 2008. **7**(11): p. 1044-55.
54. Vlaar, A., et al., *The Reliability of Transcranial Duplex Scanning in Parkinsonian Patients: Comparison of Different Observers and Ultrasound Systems.* Ultraschall Med, 2010.
55. *The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations.* Mov Disord, 2003. **18**(7): p. 738-50.
56. Gilman, S., et al., *Second consensus statement on the diagnosis of multiple system atrophy.* Neurology, 2008. **71**(9): p. 670-6.
57. Litvan, I., et al., *Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop.* Neurology, 1996. **47**(1): p. 1-9.
58. McKeith, I.G., et al., *Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium.* Neurology, 2005. **65**(12): p. 1863-72.
59. Mahapatra, R.K., et al., *Corticobasal degeneration.* Lancet Neurol, 2004. **3**(12): p. 736-43.
60. Vlaar, A.M., et al., *The predictive value of transcranial duplex sonography for the clinical diagnosis in undiagnosed parkinsonian syndromes: comparison with SPECT scans.* BMC Neurol, 2008. **8**: p. 42.

61. Bachmann, L.M., et al., *Multivariable adjustments counteract spectrum and test review bias in accuracy studies.* J Clin Epidemiol, 2009. **62**(4): p. 357-361 e2.
62. van de Loo, S., et al., *Reproducibility and diagnostic accuracy of substantia nigra sonography for the diagnosis of Parkinson's disease.* J Neurol Neurosurg Psychiatry, 2010. **81**(10): p. 1087-92.

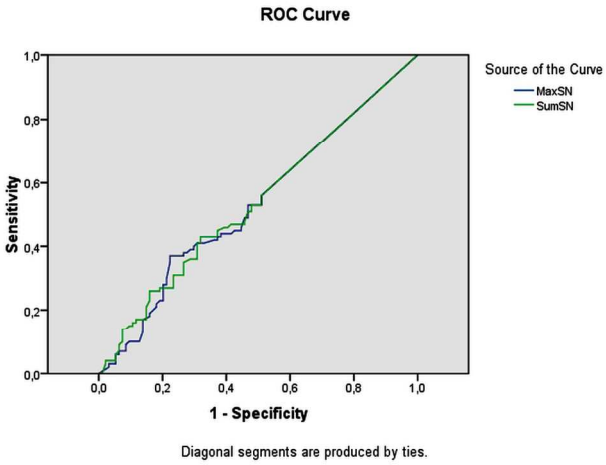
Flowchart. Trial profile:



Patiens flow chart
123x90mm (300 x 300 DPI)



Boxplot of final diagnoses compared to the range of the maximum size of the hyperechogenic substantia nigra (SN+)
160x90mm (300 x 300 DPI)



ROC Curve the maximum and the sum of the area of hyperechogenic substantia nigra (SN+) correlated with the final diagnosis Idiopathic Parkinson's Disease (IPD)
160x90mm (300 x 300 DPI)

STARD checklist for reporting of studies of diagnostic accuracy

(version January 2003)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	6
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	7
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	7
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	7
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	Prospective , see protcol
Test methods	7	The reference standard and its rationale.	9
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	9
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	9
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	9
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	9
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	10
	13	Methods for calculating test reproducibility, if done.	Ref. 54
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	7
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	7 and 12
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	Fig 1.
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	7-8
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	12
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	12-14
	20	Any adverse events from performing the index tests or the reference standard.	NA
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	Fig. 2
	22	How indeterminate results, missing data and outliers of the index tests were handled.	8-10
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	Ref. 54
	24	Estimates of test reproducibility, if done.	Ref.54
DISCUSSION	25	Discuss the clinical applicability of the study findings.	17-18