



HAL
open science

Reproducibility and diagnostic accuracy of substantia nigra sonography for the diagnosis of Parkinson's disease

Simone van de Loo, Uwe Walter, Stefanie Behnke, Johann Hagenah, Matthias Lorenz, Matthias Sitzer, Rüdiger Hilker, Daniela Berg

► **To cite this version:**

Simone van de Loo, Uwe Walter, Stefanie Behnke, Johann Hagenah, Matthias Lorenz, et al.. Reproducibility and diagnostic accuracy of substantia nigra sonography for the diagnosis of Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 2010, 81 (10), pp.1087. <10.1136/jnnp.2009.196352>. <hal-00557430>

HAL Id: hal-00557430

<https://hal.science/hal-00557430v1>

Submitted on 19 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Reproducibility and diagnostic accuracy of substantia nigra sonography for the diagnosis of Parkinson's disease

Simone van de Loo¹, Uwe Walter², Stefanie Behnke³, Johann Hagenah⁴, Matthias Lorenz¹,
Matthias Sitzer¹, Rüdiger Hilker¹⁺, Daniela Berg⁵⁺

1 Department of Neurology, University Hospital, Goethe-University,
Frankfurt am Main, Germany

2 Department of Neurology, University of Rostock, Germany

3 Department of Neurology, University of the Saarland, Homburg/Saar, Germany

4 Department of Neurology, University of Lübeck, Germany

5 Hertie Institute for Clinical Brain Research and Institute for Medical Genetics,
University of Tübingen, Germany

+ the last two authors contributed equally to this work

Key Words: Parkinson's disease, sonography, substantia nigra

Word count: **3233**

To whom correspondence should be addressed:

Ruediger Hilker, MD
Department of Neurology
Goethe University
Schleusenweg 2-16
60528 Frankfurt am Main, Germany
Phone: +49-69-6301 5852
Fax: +49-69-6301 4498
Email: hilker@med.uni-frankfurt.de

Abstract

Objective: Transcranial sonography (TCS) shows characteristic hyperechogenicity of the substantia nigra (SN) in patients with Parkinson's disease (PD). Although this feature is meanwhile well established, sufficient observer reliability and diagnostic accuracy is a prerequisite for advancements of this method. Methods: We investigated both aspects in a cross-sectional study with four blinded TCS raters in 22 PD patients and 10 healthy controls. Results: As expected, we found significant bilateral SN hyperechogenicity in PD patients. Quantitative computerized SN planimetry had a substantial intra- (ICC 0.97 and 0.93 respectively for both hemispheres) and inter-rater reliability (ICC 0.84 and 0.89), while visual semi-quantitative echogenicity grading of the SN revealed a moderate intra-rater (weighted kappa 0.80 ipsilateral and 0.74 contralateral) and slight (0.33) to fair (0.51) inter-rater reliability only. Diagnostic accuracy measured as the area-under curve (AUC) of ROC plots was highest in TCS of the SN opposite to the clinically most affected body side (planimetry 0.821, echogenicity grading 0.792) with a hyperechogenic area of 0.24 cm² as the optimum cut-off value for the differentiation between PD and controls (sensitivity 79%, specificity 81%). Conclusions: The data demonstrate that the observer variability of SN planimetry is low in the hands of experienced investigators. This approach also offers adequate diagnostic accuracy. We conclude that reliable SN TCS data on PD can be achieved in clinical routine and multi-center trials when standardized analysis protocols and certain quality criteria of brain parenchyma sonography are met.

Introduction

Transcranial B-mode sonography (TCS) of the substantia nigra (SN) is quick, safe, non-invasive and easily applicable in patients with Parkinson's disease (PD).[1, 2] TCS detects the reflection of ultrasound waves which depends among other factors on their propagation velocity in brain tissue and on tissue impedance. The characteristic finding in PD patients is a hyperechogenic SN with an enlarged area of bright ultrasound reflexes in the midbrain,[2] which presumably results from an increased tissue iron content.[3-6] The quantification of the SN signal is usually carried out by planimetric measurements of the echogenic area or by semi-quantitative visual grading.[1, 7]

Although a large body of evidence for the characteristic SN hyperechogenicity in PD is available by now, the widespread use of brain parenchyma sonography in routine diagnostics is still limited due to the frequent belief that ultrasound findings are highly subjective, examiner-dependent and of uncertain diagnostic value for these reasons. In fact, several factors are suitable to impair the inter- and intra-observer reliability of TCS, of which different experience in the evaluation of SN ultrasound images, technical factors such as image optimization by means of the ultrasonic device settings and biological conditions like the permeability of the pre-auricular acoustic bone window are the most important ones.[1, 8] Aware of these debated factors, we performed this examiner-blinded cross-sectional study on the reproducibility and diagnostic accuracy of SN TCS with the aim to provide **supportive arguments** for the routine application of this method in the diagnostic work-up of PD.

Materials and methods

Study subjects and design

After obtaining written informed consent and permission of the local ethical committee, a total of 32 individuals were included in the study: 22 patients with the diagnosis of PD according to the UK Brain Bank criteria (14 males, 8 females, mean age 65.4 ± 10.5 years, disease duration 8.0 ± 5.5 years, UPDRS motor score 18.4 ± 8.0 ; 12 akinetic-rigid, 9 equivalence, 1 tremor-dominant PD type, distribution of HY stages: I: 5, II: 10; III: 7; levodopa equivalence dose: 454 ± 168 mg) and 10 healthy age-matched controls (3 males, 7 females, mean age 63.5 ± 8.2 years). PD patients were consecutively recruited from the Movement Disorder outpatient clinic of the Department of Neurology of Goethe University Hospital Frankfurt/Main (Germany). Healthy controls were patient spouses or suitable employees of the clinic. All controls had a normal neurological status and were screened for the absence of parkinsonian motor signs by an experienced movement disorders specialist (R.H.). All individuals were screened for a sufficient pre-auricular bone window before study inclusion.

Four experienced neurologists with long-standing experience in the application of SN TCS in the university hospital setting served as study raters (Rostock: U.W., Homburg/Saar: S.B., Lübeck: J.H., Tübingen: D.B.). The study took place on two consecutive days in the Department of Neurology of Goethe University Frankfurt/Main (Germany). Per day, 16 participants were investigated twice by each rater resulting in a total of 128 scans per day and 256 scans over the entire study period. The study participants and raters were consecutively assigned to four separate shaded examination rooms in random order. Periodic changes of subjects and raters between the examination rooms were controlled and synchronized. **For best possible rater blinding, the study subjects always entered the examination room prior to investigators and awaited the TCS scan laying in supine position on the examination couch completely hidden for the later arriving**

sonographer by a large, non-transparent drapery. The investigators could not see any part of the study participants' body which might have shown parkinsonian signs (e.g. face, trunk, limbs). Examiners and study participants were not allowed to speak with each other. During TCS, the examiners were assisted by one support person who documented the data and supervised that the appointed examination settings were maintained. Each scan had to be completed within maximum 10 minutes. The raters did not have access to the results from any other TCS session. The minimum time span between the first and the second investigation of one study participant by the same rater was about one hour.

Transcranial brain parenchyma sonography (TCS)

SN TCS was performed with four machines of the Sonoline Antares system (Siemens AG, Germany) identical in construction and presetted with the same standard parameters (penetration depth 16 cm, dynamic range 55 dB, medium contour amplification). Image brightness and time gain compensation were individually adapted as needed. **In each study participant**, TCS was performed through the pre-auricular acoustic bone window according to a standard approach [1] using a 2.5 MHz transducer to visualize the midbrain in B-mode as the characteristic butterfly shaped structure at axial scanning planes (Figure 1). At first, the quality of the bone window was rated as moderate (grade 1) or good (grade 2). Afterwards, the SN on the same side as the TCS probe was identified and the patchy or band-shaped bright echogenic area was outlined after freezing and magnification of the image (SN planimetry in cm^2). Moreover, SN echogenicity was graded semi-quantitatively by the raters considering the size and the contrast of the echogenic area related to the surrounding midbrain tissue as follows: grade I = normal, grade II = moderate and grade III = markedly hyperechogenic. The SN findings were allocated to the body side with the most prominent clinical motor symptoms as ipsilateral (same side) and contralateral (opposite side). Since there is no affected side in healthy controls, we defined the left hemisphere as the ipsilateral

side in healthy subjects a-priori. To avoid any misunderstanding, it is emphasized that the terms ipsi- and contralateral in this paper always refer to the body side with predominant PD motor symptoms and do not apply to the hemispheres in relation to the TCS probe. After tilting the ultrasound probe about 10-20° upwards, the transverse diameter of the third ventricle was additionally determined since it is a well accessible comparative parameter for reliability measurements.

Statistical analysis

All statistical analyses were performed with SAS (Version 9.1, The SAS Institute, Cary, USA). To account for a potential bias of the data by bone window quality, most statistical analyses were repeated after eliminating measurements through an examiner-rated poor bone window. At first, standard descriptive statistics were applied and normal distribution of continuous parameters was assessed with QQ plots. For univariate between-group comparisons, we used Fisher's exact test for binary variables, Chi-square (χ^2) homogeneity test for nominal or ordinal variables with more than two values, and Student's *t* test for continuous variables. In multivariate analyses, models with generalized estimating equations (GEE) accounted for the interdependence of measurements. Subjects were set as identifier for repeated measurements with an independent correlation structure. The preconditions of the multivariate models were assessed with standard model diagnostics.

The inter- and intra-observer agreement of TCS parameters was calculated with weighted kappa values (Cicchetti-Allison type weights) for ordinal variables and intraclass correlation coefficients (ICC) for continuous variables. The ICC is a measure of the proportion of variance that is attributable to the objects of measurement, and has emerged as a widely accepted reliability index.[9] Due to missing values, the ICCs were calculated manually from the mean square estimates of generalized linear models accounting for the repeated measurement correlation structure. For the measurement of intra-observer reliability, we used ICC(3;k) according to Shrout and Fleiss (identical with Cronbach's alpha) to model

repeated measurements as fixed effects. For the measurement of inter-observer reliability, we used ICC(2;k) to model the raters as random effects. According to a proposal by Shrout, the following ranges of reliability values were classified: 0.00 – 0.10: virtually none; 0.11-0.40: slight; 0.41-0.60: fair; 0.61 – 0.80: moderate; 0.81-1.00: substantial.[10]

We calculated the sensitivity and specificity of TCS findings for the diagnosis of PD including testing for independence with Fisher's exact test. The discriminative power of SN TCS accounting for both, sensitivity and specificity was expressed as the area under the receiver-operating characteristic (ROC) curve (ROC-AUC). The Youden index ($J = \text{sensitivity} + \text{specificity} - 1$) provided a single numeric value with maximum of 1.00 as a global measure of test accuracy.[11, 12]

Results

Bone window quality

In 142 out of 256 measurements (56%), the examiners rated a good bone window quality (grade 2). A poor but sufficient bone window quality at least on one side (median grading 1 over all raters) was detected in 8 /22 PD patients (36%) and in 2/10 controls (20%). A poor bone window quality was significantly more frequent in study participants older than 68 years ($p = 0.004$, unpaired t test), gender had no significant influence (41% in males vs. 33% in females, $p = 0.7$, Fisher's Exact test).

Between-group comparisons: SN echogenicity and planimetry

A markedly increased echogenicity (median grading III over all raters) was detected ipsi- and contralateral to the clinically most affected body side in 16/22 PD patients (73%) which was significantly more frequent than the rate of median SN echogenicity grade III over all raters in 2/10 control subjects (20%) ($p = 0.02$ ipsilateral and $p = 0.009$ contralateral, χ^2 test). In a multivariate model adjusted for age, gender and rater, the presence of SN echogenicity grade III versus I was related with a significantly higher probability for PD with an odds ratio (OR) of 6.30 (CI95% 2.67 – 14.84; $p < 0.0001$). In contrast, echogenicity grade II versus I was not significant (OR 1.76, CI95% 0.49 – 6.27, $p = 0.4$). SN planimetry revealed significantly higher values in PD patients compared to controls with $0.26 \pm 0.05 \text{ cm}^2$ versus $0.19 \pm 0.08 \text{ cm}^2$ ipsilateral ($p = 0.01$, unpaired t test) and $0.27 \pm 0.05 \text{ cm}^2$ versus $0.19 \pm 0.06 \text{ cm}^2$ contralateral ($p = 0.0002$, unpaired t test).

Reproducibility of SN findings

Considering all measurements, the intra-observer reliability of SN planimetry ranged rater-dependent from moderate (ICC 0.75) to substantial (ICC 0.95), which was slightly below the reliability of the diameter assessments of the third ventricle (Table 1).

Table 1

Intra-observer reliability of SN planimetry, SN echogenicity grading and diameter measurement of the third ventricle presented separately for each rater and averaged for all investigators. Data are given for all measurements (n = 256) and in brackets for scans through good bone windows (n = 142).

rater	SN planimetry <i>ICC (Cronbach's alpha)</i>		SN echogenicity <i>weighted kappa</i>		Third ventricle <i>ICC (Cronbach's alpha)</i>
	ipsilateral	contralateral	ipsilateral	contralateral	
1	0.93 (0.97)	0.95 (0.95)	0.72 (0.90)	0.73 (0.79)	0.95 (0.99)
2	0.89 (0.95)	0.82 (0.91)	0.54 (0.69)	0.59 (0.53)	0.96 (0.94)
3	0.90 (0.93)	0.90 (0.92)	0.80 (0.94)	0.83 (0.94)	0.97 (0.97)
4	0.87 (0.89)	0.75 (0.92)	0.67 (0.68)	0.55 (0.68)	0.94 (0.94)
all raters	0.97 (0.99)	0.93 (0.93)	0.69 (0.80)	0.67 (0.74)	0.96 (0.96)

The ICC averaged over all raters amounted to **0.97** (ipsilateral) and 0.93 (contralateral). Similar ICC values were found in the subgroup of measurements through a good bone window. The weighted kappa values of echogenicity grading ranged investigator-dependent from fair (**0.54**) to substantial (**0.83**) in all measurements with a moderate congruency over all raters of **0.69 (ipsilateral) and 0.67 (contralateral)**. In measurements through a good bone window, the weighted kappa values were between **0.53** and 0.94 with a moderate congruency over all raters of **0.80 (ipsilateral) and 0.74 (contralateral)**. The inter-observer ICC of SN planimetry were substantial and reached 0.84 ipsi- and 0.89 contralateral (0.86 and 0.87 for good bone windows respectively). The inter-observer weighted kappa values of

SN echogenicity grading ranged from slight (0.33) to fair (0.51) in all measurements and from 0.35 to 0.49 in scans with a good bone window (Table 2).

Table 2

Inter-observer reliability of SN echogenicity grading (weighted kappa values). Data are given for all measurements (n = 256) and in brackets for scans through good bone windows (n = 142).

rater	1	2	3	4
1	1	0.51 (0.49)	0.40 (0.44)	0.33 (0.37)
2	0.51 (0.49)	1	0.35 (0.35)	0.46 (0.45)
3	0.40 (0.44)	0.35 (0.35)	1	0.41 (0.36)
4	0.33 (0.37)	0.46 (0.45)	0.41 (0.36)	1

Diagnostic accuracy of SN findings

When comparing the diagnostic accuracy of SN TCS analysis methods, the highest ROC-AUC were found in the contralateral SN (planimetry 0.821, echogenicity grading 0.792) with the most favorable sensitivity to specificity relation for the categories “hyperechogenic planimetry (> 0.25 cm²) (J = 0.52) and echogenicity grade III (J = 0.54) (Table 3). The subanalysis of good bone window measurements showed slightly higher J values and ROC-AUC in the ipsilateral SN compared to the analysis of the entire dataset, but no remarkable changes contralaterally. The ROC analysis of planimetry showed a hyperechogenic contralateral SN area of 0.24 cm² as the optimum cut-off value for the differentiation between PD and controls (sensitivity 79%, specificity 81%, J = 0.60; Figure 2).

Table 3 Diagnostic accuracy of SN planimetry and echogenicity grading for prediction of the disease status (PD or control). Sensitivity, specificity and area-under-curve (AUC) of the ROC analysis are presented. Data are given **(A)** for all measurements (n= 256) and **(B)** for measurements through good bone windows (n = 142). * p < 0.0001 (Fisher's exact test)

A		ipsilateral				contralateral			
		Sensitivity (%)	Specificity (%)	J	ROC-AUC	Sensitivity (%)	Specificity (%)	J	ROC-AUC
Planimetry	Increased (0.2-0.25 cm ²)	79.0	63.8	0.43	0.718 *	84.1	63.8	0.48	0.821 *
	Hyperechogenic (> 0.25 cm ²)	51.7	75.0	0.27		67.0	85.0	0.52	
Echogenicity	Grade II	85.1	43.8	0.29	0.726 *	89.7	48.8	0.39	0.792 *
	Grade III	64.9	76.3	0.41		73.7	80.0	0.54	

B		ipsilateral				Contralateral			
		Sensitivity (%)	Specificity (%)	J	ROC-AUC	Sensitivity (%)	Specificity (%)	J	ROC-AUC
Planimetry	Increased (0.2-0.25 cm ²)	87.8	60.3	0.48	0.740 *	88.1	61.8	0.50	0.833 *
	Hyperechogenic (> 0.25 cm ²)	61.2	71.4	0.33		73.3	83.6	0.57	
Echogenicity	Grade II	90.8	47.6	0.38	0.766 *	90.0	47.3	0.37	0.795 *
	Grade III	77.6	71.4	0.49		78.2	78.2	0.56	

Discussion

This study reports data on the inter- and intra-rater reproducibility and diagnostic accuracy of different SN TCS analysis approaches applied by four blinded examiners in PD patients and controls. Other than previous studies which reported simple correlation coefficients as a measure of inter-observer variability in SN TCS,[8, 13-16] we calculated established reliability indices (weighted kappa, ICC), since they do account for the random-chance of rater agreement.[17]

As expected, our data proved a significant bilateral enlargement of the hyperechogenic SN in PD patients versus controls which is in line with several previous studies.[1] The substantial intra- and inter-observer reliability of quantitative SN planimetry validates the reproducibility of this quantitative method in the diagnostic work-up of PD patients. Two previous papers reported comparably high intra- and inter-observer reliability for SN planimetry in healthy controls with ICC 0.98 [18] and Cohen's Kappa 0.83.[19] In contrast, we found a remarkably lower rater congruency for the semi-quantitative SN echogenicity grading which achieved fair (intra-observer) to moderate (inter-observer) weighted kappa values at best. This is in contrast to a recent study reporting Spearman rank correlation coefficients of $r = 0.82$ (inter-observer) and of 0.96 (intra-observer) for a five-grade SN echogenicity rating scale.[8] The apparent discrepancy between these and our own results might be the use of different statistical methods and SN grading scales with three versus five steps. Our data indicate that the enlargement of hyperechogenic signals in the SN compared to the echogenicity of the surrounding brain tissue is hard to measure reliably in a simple three-step semi-quantitative approach. Rather, they suggest that the visual echogenicity grading considering the contrast of the SN to the surrounding midbrain and the size of the hyperechogenic area is much more influenced by the subjective impression of the investigator than the encircling of the SN area yielding interval-scaled quantitative data. Therefore, we suggest preferring SN planimetry with a carefully defined cut-off value for the diagnosis of PD.

A crucial point of a diagnostic test is its accuracy in the prediction of a disease. Since the investigator may be strongly biased by the personal knowledge or the presumption of the diagnosis due to apparent clinical symptoms of the examined person, we have chosen a sonographer-blinded study design. We were able to show that a marked enlargement of the SN area in planimetry as well as the semi-quantitative grading III revealed ROC-AUC values which fall in the category of a moderately accurate diagnostic tool.[20] The latter serve as a global statistic measure of the overall test performance.[21] Moreover, the ROC analysis allowed us to calculate the optimal SN planimetry cut-off value for the distinction between PD and controls in our study cohort. We found that 0.24 cm² for the SN contralateral to the most affected body side meets these criteria with nearly 80% sensitivity and specificity likewise ($J = 0.60$), which is in line with recent findings on the sensitivity and specificity of SN TCS for the diagnosis of PD.[22, 23] **However, it is important to note that the 0.24 cm² cut-off was calculated in a relatively small study cohort and that previous studies reported slightly different values for various ultrasound systems. Most studies published so far were performed on the Sonoline Elegra system (Siemens, Erlangen, Germany) with three classification categories of the SN echogenic area, namely < 0.20 cm² normal, 0.20 - 0.25 cm² moderately and > 0.25 cm² markedly hyperechogenic.[1, 19, 24, 25] Another study using this system defined the 80th percentile of control values (0.20 cm²) as cut-off and found 0.83 as positive and 0.78 as negative predictive value for the diagnosis of PD.[26] Glaser et al. showed a nearly 1.2-fold larger echogenic SN area measured with the Sonoline Antares compared with the Elegra system in the same individuals.[27] A slightly higher optimum cut-off value of 0.27 cm² was published by Hagenah and colleagues using the SONOS 5500 machine (Philips Medical Systems, Best, The Netherlands).[28] These to a certain degree divergent data emphasize that each laboratory has to establish its own standard values for SN echogenicity, depending among other factors on the ultrasound system used.[7]**

Hyperechogenicity of the SN can be found in about 10% of asymptomatic individuals up to the age of 79 years,[3, 19] which has possible impact on the diagnostic accuracy and the specificity of the method in particular. The frequency of this SN feature fits well with the known 10% prevalence of incidental Lewy Bodies in the SN of asymptomatic persons aged over 50 years.[29] Several data demonstrated that a hyperechogenic SN in individuals without PD is related to lower presynaptic dopaminergic tracer uptake in PET and SPECT,[3, 19, 30] to motor slowness and extrapyramidal symptoms in higher age [13] and to an increased occurrence of non-motor PD symptoms, such as olfactory dysfunction or depression.[16, 30] Therefore, SN hyperechogenicity has been suggested to represent a vulnerability factor of the nigrostriatal system for neuronal degeneration and a marker of individual predisposition for the development of PD,[19] which is currently investigated in longitudinal studies.

We did not calculate the predictive value of SN TCS for PD, since this parameter is highly dependent on the estimated total prevalence of the disease in the study population. Though PD is a common neurological disorder, it has a relatively low and age-dependent prevalence in the total population of about 1-2% in elderly people over the age of 60 years.[31] In this case, the predictive value of a given diagnostic test is remarkably lower due to the high a-priori probability of false positive test results.[32] To mitigate this general problem for screening tests, it is helpful to restrict the target population to a smaller group of at-risk individuals for PD, for example to those with a positive family history and with pre-motor PD symptoms, such as olfactory dysfunction, dysautonomia or mood and sleep disorders.[33]

It is important to note that our study design included some prerequisites which constitute ideal but in real practice not always given examination conditions and, therefore, limit the generalization of our results. At first, only well-experienced sonographers took part in the study in order to avoid a loss of reliability due to a disparate training level of the investigators. A recent study by Skoloudik and colleagues clearly showed a much poorer correlation of SN

TCS results in non-trained sonographers.[8] Thus, the reliability rates found by us can only be expected in trained investigators. However, the fact that the person who performs or interprets an imaging investigation needs to be trained is self-evident and holds true for any other imaging technique. **Although explicit recommendations have not been formally defined as yet, we would like to suggest that the education level for SN ultrasound investigators should be geared to the training standards for vascular ultrasound in stroke neurologists and neuroradiologists.** It has also to be kept in mind that we exclusively used a well-established ultrasound system for depiction of the SN (Sonoline Antares, Siemens AG, Germany).[34] However, the technical development of various ultrasound systems is different and not necessarily adapted to the requirements of optimum brain parenchyma sonography. Also, our data clearly show that the observer reliability more than the diagnostic accuracy of SN TCS is dependent on the quality of the pre-auricular acoustic bone window yielding more favorable data in measurements through well permeable windows allowing adequate insonation of the midbrain. Finally, the number of patients and especially controls is small. This should especially be considered when interpreting the results concerning prediction of the disease. All these considerations underline that each examiner has to interpret the results of brain parenchyma sonography carefully in the context of the individual scanning conditions.

In conclusion, our data show that the observer variability of SN TCS is low in the hands of experienced investigators. The method offers sufficient observer reliability when the interval-scaled planimetry of the echogenic SN area is used. Adequate diagnostic accuracy of SN TCS with nearly 80% sensitivity and specificity for PD diagnosis is assured by SN planimetry with an adequate cut-off value. We conclude that reliable SN TCS data on PD can also be achieved in multi-center trials with a larger number of investigators when standardized analysis protocols and a strict quality control are implemented.

Acknowledgements

We thank the Siemens AG (Erlangen, Germany) for providing four ultrasound devices of the Sonoline Antares system. This study was funded by a grant of the Deutsche Parkinson Vereinigung e.V. (dPV).

Copyright licence statement

The corresponding author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd, and its licensees to permit this article (if accepted) to be published in the Journal of Neurology, Neurosurgery and Psychiatry and any other BMJ PGL products and to exploit all subsidiary rights, as set out in our licence.

Competing interests

There are no financial interests to disclose.

Figure legends

Figure 1 Transaxial midbrain B-mode sonography (Sonoline Antares, Siemens AG, Germany) of the substantia nigra (SN) measured twice by four raters in the same PD patient with good transtemporal bone window quality (1st scan: upper row, 2nd scan: lower row). In each case, the SN was encircled for computerized measurement (SN planimetry). The individual planimetry values were (1st scan / 2nd scan): rater 1: 0.26 / 0.32; rater 2: 0.26 / 0.29; rater 3: 0.28 / 0.28; rater 4: 0.28 / 0.27.

Figure 2 Receiver-operating characteristic (ROC) plot of contralateral SN planimetry for the prediction of PD. The area-under-curve (ROC-AUC) is 0.821. Dotted lines show the tangent of the diagonal with the ROC curve and the perpendiculars from this point (black circle) to both axes indicating 81% specificity (1-specificity = 19% on x-axis) and 79% sensitivity (on y-axis). The black circle represents the optimum cut-off value of contralateral SN planimetry (0.24 cm²) for the distinction between PD and controls (Youden Index $J = 0.60$).

References

1. Berg D, Godau J, Walter U. Transcranial sonography in movement disorders. *Lancet Neurol* 2008;7:1044-55.
2. 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;45:182-4.
3. 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.
4. Berg D. In vivo detection of iron and neuromelanin by transcranial sonography--a new approach for early detection of substantia nigra damage. *J Neural Transm* 2006;113:775-80.
5. 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;25:901-4.
6. Zecca L, Berg D, Arzberger T, Ruprecht P, Rausch WD, Musicco M, Tampellini D, Riederer P, Gerlach M, Becker G. In vivo detection of iron and neuromelanin by transcranial sonography: A new approach for early detection of substantia nigra damage. *Mov Disord* 2005;20:1278-85.
7. Walter U, Behnke S, Eyding J, Niehaus L, Postert T, Seidel G, Berg D. Transcranial brain parenchyma sonography in movement disorders: State of the art. *Ultrasound Med Biol* 2007;33:15-25.

8. Skoloudik D, Fadrna T, Bartova P, Langova K, Ressner P, Zapletalova O, Hlustik P, Herzig R, Kannovsky P. Reproducibility of sonographic measurement of the substantia nigra. *Ultrasound Med Biol* 2007;33:1347-52.
9. Shoukri MM. Reliability for continuous scale measurements. In: Shoukri MM, ed. *Measures of interobserver agreement*. Boca Raton: Chapman & Hall/CRC; 2004.
10. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res* 1998;7:301-17.
11. Reiser B. Measuring the effectiveness of diagnostic markers in the presence of measurement error through the use of roc curves. *Stat Med* 2000;19:2115-29.
12. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-5.
13. Berg D, Siefker C, Ruprecht-Dorfler P, Becker G. Relationship of substantia nigra echogenicity and motor function in elderly subjects. *Neurology* 2001;56:13-7.
14. Puls I, Berg D, Maurer M, Schliesser M, Hetzel G, Becker G. Transcranial sonography of the brain parenchyma: Comparison of b-mode imaging and tissue harmonic imaging. *Ultrasound Med Biol* 2000;26:189-94.
15. 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;21:94-8.
16. Walter U, Hoepfner J, Prudente-Morrissey L, Horowski S, Herpertz SC, Benecke R. Parkinson's disease-like midbrain sonography abnormalities are frequent in depressive disorders. *Brain* 2007;130:1799-807.
17. Downing SM. Reliability: On the reproducibility of assessment data. *Med Educ* 2004;38:1006-12.

18. Huang YW, Jeng JS, Tsai CF, Chen LL, Wu RM. Transcranial imaging of substantia nigra hyperechogenicity in a taiwanese cohort of parkinson's disease. *Mov Disord* 2007;22:550-5.
19. 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;53:1026-31.
20. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-93.
21. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000;45:23-41.
22. Gaenslen A, Unmuth B, Godau J, Liepelt I, Di Santo A, Schweitzer KJ, Gasser T, Machulla HJ, Reimold M, Marek K, Berg D. The specificity and sensitivity of transcranial ultrasound in the differential diagnosis of parkinson's disease: A prospective blinded study. *Lancet Neurol* 2008;7:417-24.
23. Prestel J, Schweitzer KJ, Hofer A, Gasser T, Berg D. Predictive value of transcranial sonography in the diagnosis of parkinson's disease. *Mov Disord* 2006;21:1763-5.
24. **Berg D, Siefker C, Becker G. Echogenicity of the substantia nigra in parkinson's disease and its relation to clinical findings. *J Neurol* 2001;248:684-9.**
25. **Walter U, Wittstock M, Benecke R, Dressler D. Substantia nigra echogenicity is normal in non-extrapyramidal cerebral disorders but increased in parkinson's disease. *J Neural Transm* 2002;109:191-6.**

26. Behnke S, Berg D, Naumann M, Becker G. Differentiation of parkinson's disease and atypical parkinsonian syndromes by transcranial ultrasound. *J Neurol Neurosurg Psychiatry* 2005;76:423-5.
27. Glaser M, Weber U, Hinrichs H, Niehaus L. Neuroimaging of midbrain structures with different phased-array ultrasound systems. *Klin Neurophysiol* 2006;37:165-8.
28. Hagenah JM, Konig IR, Becker B, Hilker R, Kasten M, Hedrich K, Pramstaller PP, Klein C, Seidel G. Substantia nigra hyperechogenicity correlates with clinical status and number of parkin mutated alleles. *J Neurol* 2007;254:1407-13.
29. Forno LS. Concentric hyalin intraneuronal inclusions of lewy type in the brains of elderly persons (50 incidental cases): Relationship to parkinsonism. *J Am Geriatr Soc* 1969;17:557-75.
30. Sommer U, Hummel T, Cormann K, Mueller A, Frasnelli J, Kropp J, Reichmann H. Detection of presymptomatic parkinson's disease: Combining smell tests, transcranial sonography, and spect. *Mov Disord* 2004;19:1196-202.
31. Lees AJ, Hardy J, Revesz T. Parkinson's disease. *Lancet* 2009;373:2055-66.
32. Galen RS. *Application of the predictive value model in the analysis of test effectiveness. Test selection strategies, clinics in laboratory medicine*. Philadelphia: Saunders; 1982.
33. Tolosa E, Gaig C, Santamaria J, Compta Y. Diagnosis and the premotor phase of parkinson disease. *Neurology* 2009;72:S12-20.

34. Walter U, Kanowski M, Kaufmann J, Grossmann A, Benecke R, Niehaus L.
Contemporary ultrasound systems allow high-resolution transcranial imaging
of small echogenic deep intracranial structures similarly as MRI: A phantom
study. *Neuroimage* 2008;40:551-8.

Figure 1

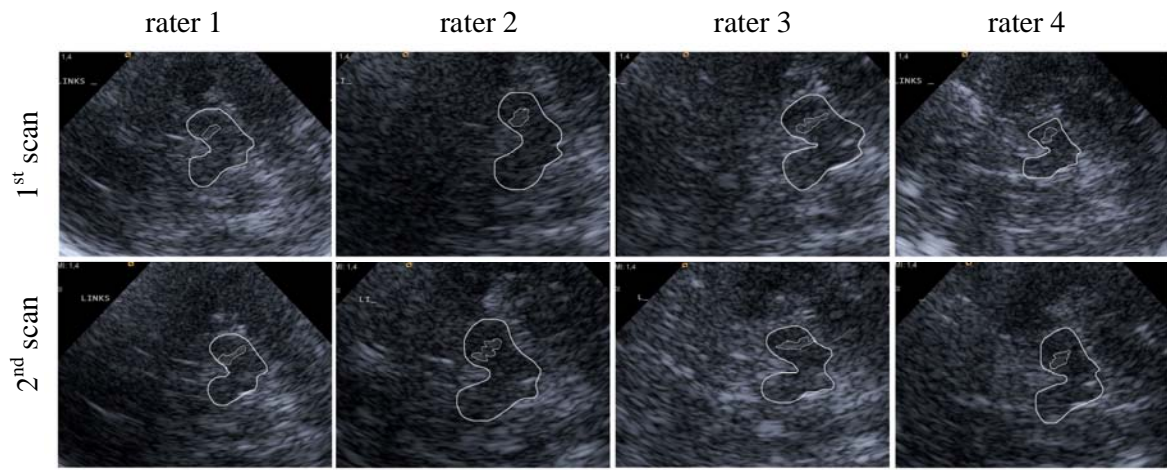


Figure 2

