

Composite cerebellar functional severity score: validation of a quantitative score of cerebellar impairment

Sophie Tezenas du Montcel,^{1,2} Perrine Charles,^{3,4,5,6} Pascale Ribai,^{3,4,5} Cyril Goizet,^{4,5,7} Alice Le Bayon,⁸ Pierre Labauge,⁸ Lucie Guyant-Maréchal,⁹ Sylvie Forlani,^{4,5} Celine Jauffret,³ Nadia Vandenberghe,¹⁰ Karine N'Guyen,¹¹ Isabelle Le Ber,^{4,5,6} David Devos,¹² Carlo-Maria Vincitorio,³ Mario-Ubaldo Manto,¹³ François Tison,⁷ Didier Hannequin,⁹ Merle Ruberg,^{3,4,5} Alexis Brice^{3,4,5,6,14} and Alexandra Dürr^{3,4,5}

¹AP-HP, Department of Biostatistics and Medical Informatics and Pitié-Salpêtrière Charles-Foix Clinical Research Unit, ²University Pierre et Marie Curie, EA 3974, Modelling in Clinical Research, ³AP-HP, Pitié-Salpêtrière Hospital, Department of Genetics and Cytogenetics, ⁴INSERM, UMR_S679 Neurologie & Thérapeutique Expérimentale, ⁵Federative Institute for Neuroscience Research (IFR70), Pitié-Salpêtrière Hospital, ⁶AP-HP, Pitié-Salpêtrière Hospital, Federation of Neurology, F-75013 Paris, ⁷Bordeaux CHU Hospitals, Pellegrin Hospital, Department of Medical Genetics and Federation of Clinical Neurosciences, Bordeaux, ⁸Nîmes Hospitals, Carémeau Hospital, Service of Neurology, Nîmes, ⁹Rouen Hospitals, Charles Nicolle Hospital, Clinical Genetics Unit, Rouen, ¹⁰Lyon Hospitals, Pierre Wertheimer Neurological Hospital, Service of Neurology C, Lyon, ¹¹Marseille Hospitals, La Timone Hospital, Department of Genetics, Marseille, ¹²Lille CHU Hospitals, Roger Salengro Hospital, Service of Neurology, Lille, France, ¹³FNRS, ULB Erasme, Service of Neurology, Brussels, Belgium and ¹⁴UPMC Univ Paris 06, UMR_S679, F-75005, Paris, France

Correspondence to: Alexandra Dürr, MD, PhD, INSERM UMR_S679 and Department of Genetics and Cytogenetics, Hôpital Pitié-Salpêtrière, 47-83 boulevard de l'Hopital, 75651 Paris Cedex 13- France.

E-mail: durr@ccr.jussieu.fr

Reliable and easy to perform functional scales are a prerequisite for future therapeutic trials in cerebellar ataxias. In order to assess the specificity of quantitative functional tests of cerebellar dysfunction, we investigated 123 controls, 141 patients with an autosomal dominant cerebellar ataxia (ADCA) and 53 patients with autosomal dominant spastic paraplegia (ADSP). We evaluated four different functional tests (nine-hole pegboard, click, tapping and writing tests), in correlation with the scale for the assessment and rating of cerebellar ataxia (SARA), the scale of functional disability on daily activities (part IV of the Huntington disease rating scale), depression (the Public Health Questionnaire PHQ-9) and the EQ-5D visual analogue scale for self-evaluation of health status. There was a significant correlation between each functional test and a lower limb score. The performance of controls on the functional tests was significantly correlated with age. Subsequent analyses were therefore adjusted for this factor. The performances of ADCA patients on the different tests were significantly worse than that of controls and ADSP patients; there was no difference between ADSP patients and controls. Linear regression analysis showed that only two independent tests, the nine-hole pegboard and the click test on the dominant side ($P < 0.0001$), accounted for the severity of the cerebellar syndrome as reflected by the SARA scores, and could be represented by a composite cerebellar functional severity (CCFS) score calculated as follows:

$$\text{CCFS} = \log_{10} \left(7 + \frac{\text{Z pegboard dominant hand}}{10} + 4 \times \frac{\text{Z click dominant hand}}{10} \right)$$

The CCFS score was significantly higher in ADCA patients compared to controls (1.12 ± 0.18 versus 0.85 ± 0.05 , $P_c < 0.0001$) and ADSP patients (1.12 ± 0.18 versus 0.90 ± 0.08 , $P_c < 0.0001$) and was correlated with disease duration ($P < 0.0001$) but independent of self-evaluated depressive mood in ADCA. Among genetically homogeneous subgroups of ADCA patients (Spinocerebellar ataxia 1, 2, 3), SCA3 patients had significantly lower (better) CCFS scores than SCA2 ($P_c < 0.04$) and the same tendency was observed in SCA1. Their CCFS scores remained significantly worse than those of ADSP patients with identified SPG4 mutations ($P < 0.0001$). The pegboard and click tests are easy to perform and accurately reflect the severity of the disease. The CCFS is a simple and validated method for assessing cerebellar ataxia over a wide range of severity, and will be particularly useful for

discriminating paucisymptomatic carriers from affected patients and for evaluating disease progression in future therapeutic trials.

Keywords: cerebellar ataxia; spastic paraplegia; Composite Cerebellar Functional Severity (CCFS) score; natural history

Abbreviations: ADCA = autosomal dominant cerebellar ataxia; ADSP = autosomal dominant spastic paraplegia; CCFS = composite cerebellar functional severity; ICARS = International Cooperative Ataxia Rating Scale; MSFC = multiple sclerosis functional composite; PASAT = paced auditory serial addition test; SARA = Scale for the Assessment and Rating of Ataxia; VAS = visual analogue scale.

Received October 26, 2007. Revised February 20, 2008. Accepted February 25, 2008

Introduction

Autosomal dominant cerebellar ataxias (ADCA) are a group of neurodegenerative disorders that are clinically and genetically heterogeneous, with at least 29 loci/genes designated SCA (spinocerebellar ataxia) (Harding, 1993). They are characterized by core symptoms related to cerebellar dysfunction, such as ataxic gait, cerebellar dysarthria, intentional tremor and hypermetria in the limbs. Structures outside the cerebellum are also involved, and vary according to the genotype: a pyramidal syndrome is predominant in SCA1; dystonia, parkinsonism and peripheral neuropathy are frequent in SCA3; slow saccades are associated with SCA2 (Schols *et al.*, 1997; Stevanin *et al.*, 2002). Abnormal polyglutamine expansions are the underlying pathological mechanism in seven forms of the disease: SCA1, 2, 3, 6, 7, 17 and Dentatorubropallidolysian atrophy DRPLA (Orr *et al.*, 1993; Kawaguchi *et al.*, 1994; Koide *et al.*, 1994; Nagafuchi *et al.*, 1994; Imbert *et al.*, 1996; Sanpei *et al.*, 1996; David *et al.*, 1997; Zhuchenko *et al.*, 1997). These seven mutations are tested routinely in diagnostic laboratories and are responsible for 40–80% of ADCAs. The relative frequencies of these diseases vary according to geographical origin; SCA3 is the most common genotype worldwide, followed by SCA1, 2, 6, 7 and 17 (Stevanin *et al.*, 2002).

Since these disorders progress slowly, the evaluation of potential treatments that might slow their course requires reliable and easy to use tools that can assess small clinical changes over short periods of time. Several scales are already used to evaluate the severity of cerebellar ataxia, but all have some drawbacks. The lengthy International Cooperative Ataxia Rating Scale (ICARS), a 100-point scale (Trouillas *et al.*, 1997), has been shown to have a ceiling effect, and is, therefore, not usable when linear results are needed (Schmitz-Hübsch *et al.*, 2006a). A new Scale for the Assessment and Rating of Ataxia (SARA) has been validated in dominant ataxias as part of a European effort (EUROSCA) (Schmitz-Hübsch *et al.*, 2006b), and includes three measures of posture and gait, one measure of speech and four of kinetic dysfunction. The scale is semi-quantitative, however, and relies on subjective ratings by clinicians. For patients with autosomal dominant spastic paraplegia (ADSP), the Ashworth scale is generally used to

assess spasticity in the lower limbs, but ignores possible upper limb involvement. Quantitative scales have also been developed to assess other neurological conditions. The multiple sclerosis functional composite (MSFC) score, consisting of quantitative tests of arm and hand function (nine-hole peg test), cognitive function (paced auditory serial addition test, PASAT), leg function and ambulation (TWT) (Cutter *et al.*, 1999), was developed to evaluate the major clinical manifestations of cerebellar dysfunction with non-redundant parameters that are simple to measure, sensitive to change and useful for clinical trials (Cutter *et al.*, 1999). A Friedreich ataxia composite score was developed from an arm and hand function test (9-HPT), a leg function and ambulation test (T25WT) and a speech test (PATA) (Lynch *et al.*, 2006).

The aims of the present study were to: (i) evaluate four quantitative tests of cerebellar signs in the upper limbs; (ii) find correlations with relevant clinical signs; (iii) develop a simple composite functional score that reflects these signs and that can be used in future clinical trials. To determine the specificity for ADCA of the functional tests of cerebellar dysfunction, we compared ADCA patients with a group of control subjects and with patients with so-called pure forms of ADSP that are usually limited to pyramidal signs in the lower limbs caused by degeneration of the corticospinal tract; unlike cerebellar ataxias, upper limb involvement is rare (Depienne *et al.*, 2007).

Methods

Patients and controls

Patients were included if they: (i) had a family history compatible with dominant transmission; (ii) were over 18 years of age; and (iii) had cerebellar ataxia (ADCA) or spastic paraplegia (ADSP). The ADCA group included 141 patients with ADCA (54% women): 22 SCA1, 32 SCA2, 51 SCA3, 5 SCA6, 8 SCA7, 1 SCA14, 1 SCA25, 1 SCA28 and 20 unknown SCA. The ADSP group included 53 patients with ADSP (38% women): 22 SPG4, 4 SPG3, and 27 unknown SPG.

The protocol was approved by the local ethics committee (University Hospital Pitié-Salpêtrière, Paris, France). All patients gave their written informed consent. They were seen in an out-patient setting of each university hospital, located in Paris ($n=144$), Bordeaux ($n=24$), Nîmes ($n=10$), Rouen ($n=8$), Lyon ($n=6$) and Marseille ($n=2$).

This study also includes 123 healthy volunteers (54% women) over 18 years of age. For each 10-year-class and for each sex, we included unrelated healthy individuals recruited at the out-patient clinic, without neurological or osteo-articular pathology (ages 20–29 years $n=25$; 30–39 $n=23$, 40–49 $n=27$, 50–59 $n=21$, 60–69 $n=18$, 70–79 $n=9$).

Clinical evaluations

The patients were evaluated with the following already validated and commonly used rating scales.

Clinical rating scale

The SARA was used; it includes eight items, for a total score of 0 (best) to 40 (worse) (Schmitz-Hübsch *et al.*, 2006b).

Assessment of functional disability

The functional assessment part of the Unified Huntington's Disease Rating Scale (UHDRS part IV) (Huntington Study Group, 1996) was used; it includes questions related to the performance of daily activities such as shopping, cleaning, dressing, etc., for a total score of 0 (worse) to 25 (best).

Health-related quality of life (EQ-5D)

This is a generic instrument, developed and validated by the EuroQoL Group (1990), in which the patient self-rates in particular his/her health status on a visual analogue scale (VAS); the best score is 100.

Patient health questionnaire (PHQ-9)

Patients self-evaluate the severity and the frequency of nine situations related to depression during the last 2 weeks. The method regroups the items and scores depression as major, mild to moderate or no depressive syndrome, according to Spitzer *et al.*, 1999.

Functional tests

Writing test

The patient is asked to write a standard sentence ('*maître corbeau sur un arbre perché*'), with his dominant hand, as fast as possible, but legibly. He is timed from when he begins to write until he completes the sentence.

Tapping test

This pronation and supination test evaluates hand diadochokinaesia. The patient is comfortably seated and is asked to rapidly alternate pronation and supination of the hand on his (her) thigh ten times. The patient is instructed to perform the movements as fast as possible, but to be sure to clearly hit the thigh with the back of the hand and the palm and to avoid counting the cycles mentally in order to concentrate on the movements. The trial is performed once each for the dominant and non-dominant hand. Timing begins when the back of the hand first hits the thigh and ends the 10th time the palm hits the thigh.

Nine-hole pegboard test

The patient, who is seated, holds nine dowels (9 mm in diameter and 32-mm long) in one hand and places them randomly, one

by one, with the other hand in a board with nine holes (Mathiowetz *et al.*, 1985). Timing begins when the first peg is placed in a hole and ends when the last peg is placed. The examiner holds the board steady on the table during the test. One trial is performed with each hand. If the patient drops a peg the examiner stops the timer and the patient starts the test again once from the beginning.

Click test

This test measures specifically finger-pointing coordination. It uses a simple homemade device composed of two mechanical counters fixed on a wooden board 39 cm apart. This distance of 39 cm is based on kinematic and kinetic analysis showing that it is the optimal inter-counter distance to assess the metrics of goal-directed visually guided multi-joint movements in upper limbs (Manto M-U., unpublished data). The patient is seated facing the examiner across a table on which the counters are placed (numbers facing the examiner). The patient uses his index finger to press the buttons on the counters alternately 10 times. Timing begins when the first button is pressed and stops when the second counter reaches 10. The trial is performed once with each hand.

Lower limb score

The walking distance without aid during 5 s was categorized in six subgroups: >5 m, >4 m, >3 m, >2 m, >1 m, <1 m, ranging from 0 to 6, respectively. This test was performed only in patients who were able to walk.

Statistical analyses

Test performances (time values) of the dominant hand and the non-dominant hand were compared using the Wilcoxon signed rank sum test. The comparison of performances for different functional tests was done according to age groups and sex using the Kruskal–Wallis Test. Polynomial regression (linear, quadratic and cubic factors) was performed to test the influence of ageing on the functional tests. The P -values were corrected for multiple comparisons using Bonferroni correction.

The clinical characteristics of the ADSP and ADCA patients were compared by Student's t -test or Chi-square depending on the variables.

ANOVA was used to compare performances between controls and patients. Tukey–Kramer adjustments of significance levels were used for *post hoc* comparisons (P_c).

Univariate and multivariate linear regression modelling was performed to explain disease severity (measured by SARA) by the functional tests. For the multivariate regression, a stepwise selection procedure was used. All confounding factors such as sex, disease duration, daily life activities, depression and quality of life were tested on the composite score. All tests were two-sided and P -values <0.05 were considered significant. Statistical analyses were performed using the SAS 9.1 statistical package (SAS Institute, Cary, NC, USA).

Results

Controls (Table I)

The performances of controls on all of the tests are given in Table 1. Scores on the tapping test with the dominant and

non-dominant hand were similar, with the same mean value and nearly the same standard deviations. This allowed us to use a single value to calculate the Z-scores, the mean between the dominant and non-dominant hand. In contrast, the mean values on the dominant side were significantly shorter on the pegboard ($P < 0.0001$) and click ($P < 0.0001$) tests. There were no differences according to sex. However, the mean values on each test increased significantly with age. A linear model fitted all functional tests except for the pegboard where a linear plus quadratic model fitted the data (Table 1 and Fig. 1, Supplementary data). Therefore, to eliminate age effect on the tests, in subsequent analyses, the values were adjusted by age group. Both-sided test Z-scores were calculated for both the dominant and the non-dominant side. For the tapping test, as no differences were shown in controls between dominant and non-dominant hand, a common adjustment was used. The Z-score was obtained by subtracting the expected time obtained with the formulae in Table 1 from the observed time. For instance, for the click test the Z-score is obtained using the following adjustment:

$$Z \text{ click Dominant} = \text{Observed click Dominant hand} - (8.0 + 0.05 \times \text{Age at examination}).$$

Patients

Clinical characteristics (Table 2)

Forty-nine per cent of the 194 ADCA and ADSP patients were women. There were significantly more men in the ADSP group than in the ADCA group (62% versus 46%, $P < 0.05$). The two groups of patients did not differ in age at examination (ADCA: 47.4 ± 13.4 years versus ADSP: 48.2 ± 14.2 years). ADSP patients had a significantly earlier age at onset (22.8 ± 17.4 years versus 35.8 ± 12.3 years; $P < 0.0001$), and therefore longer disease durations than those with ADCA (24.9 ± 17.3 years versus 11.6 ± 7.7 years; $P < 0.0001$).

SARA values ranged from 1 to 39 in ADCA patients, with a mean of 14.6 ± 8.2 . As expected, ADSP patients had a significantly lower (better) mean SARA score than patients with ADCA (5.8 ± 5.8 versus 14.6 ± 8.2 ; $P < 0.0001$), but slightly higher (worse) scores than normal controls (0.4 ± 1.1 , range 0 to 7.5) (Schmitz-Hübsch *et al.*, 2006a, b). Nineteen per cent of the ADSP patients had SARA scores above 7.5. ADCA patients were functionally more impaired on UHDRS scoring (22.6 ± 4.3 versus 18.0 ± 7.1 ; $P < 0.0001$), but did not differ from ADSP patients in terms of self-evaluated health status or depression. Self-evaluated depressive mood was frequent in both groups.

Performance on the functional tests

There was a significant correlation between each functional test and the lower limb score, with Pearson correlation coefficients ranging from 0.43 (writing, $P < 0.0001$) to 0.56 (click non-dominant hand, $P < 0.0001$). However, the lower

Table 1 Times of 123 healthy controls on the four functional tests according to age

Test	Time (seconds)						Polynomial model	
	Age classes (Mean \pm SD)							
	All ages	20–29 years n = 25	30–39 years n = 23	40–49 years n = 27	50–59 years n = 21	60–69 years n = 18		70–79 years n = 9
Tapping D	6.11 \pm 1.29	5.64 \pm 0.81	5.87 \pm 1.01	6 \pm 1.3	6.14 \pm 1.24	6.72 \pm 1.53	7.11 \pm 1.9	
Tapping non-D	6.11 \pm 1.21	5.44 \pm 0.77	5.83 \pm 0.83	5.96 \pm 1.13	6.24 \pm 1.3	6.83 \pm 1.38	7.44 \pm 1.24	
Tapping ^a	6.11 \pm 1.18	5.54 \pm 0.69	5.85 \pm 0.83	5.98 \pm 1.14	6.19 \pm 1.22	6.78 \pm 1.35	7.28 \pm 1.52	0.0012
Pegboard D	10.74 \pm 1.57	10.72 \pm 1.54	10.3 \pm 1.43	10.26 \pm 1.53	10.9 \pm 1.34	10.94 \pm 1.63	12.67 \pm 1.22	0.0114
Pegboard non-D	12.07 \pm 1.85	11.48 \pm 1.33	11.65 \pm 1.67	11.41 \pm 1.53	12.24 \pm 1.67	13 \pm 1.33	14.56 \pm 2.96	0.0012
Click D	10.36 \pm 1.89	9.44 \pm 1.08	9.61 \pm 1.44	10.11 \pm 1.55	10.95 \pm 2.29	11.44 \pm 1.98	12 \pm 2.18	0.0012
Click non-D	10.85 \pm 1.93	9.96 \pm 0.84	10.48 \pm 1.65	10.56 \pm 1.67	10.95 \pm 2.18	11.94 \pm 2.13	12.67 \pm 2.65	0.0414
Writing D	10.80 \pm 2.47	9.88 \pm 1.9	10.13 \pm 1.94	10.3 \pm 2.22	11.38 \pm 2.94	12.72 \pm 2.61	11.44 \pm 2.3	0.0126

D = dominant hand; nonD = non-dominant hand; P_c = P-value after Bonferroni correction for the six comparisons.

^aIdentical results were obtained with the dominant and non-dominant hand in this test. Only the mean of the two hands was used to compare the effects of age.

Table 2 Clinical characteristics of patients with autosomal dominant cerebellar ataxias (ADCA) or autosomal dominant spastic paraplegia (ADSP)

Characteristics	ADCA <i>n</i> = 141	ADSP <i>n</i> = 53	<i>P</i> -value
Male/Female ^a	65 (46%)/76 (54%)	33 (62%)/20 (38%)	<0.05
Age in years ^b	47.3 ± 13.4 [15–83]	48.2 ± 14.2 [16–80]	>0.69
Age at onset in years (<i>n</i> = 188) ^b	35.8 ± 12.3 [5–67]	22.8 ± 17.4 [0–56]	<0.0001
Disease duration in years (<i>n</i> = 188) ^b	11.6 ± 7.7 [0–44]	24.9 ± 17.3 [2–69]	<0.0001
SARA scores (/40 worse) (<i>n</i> = 193) ^b	14.6 ± 8.2 [1–39]	5.8 ± 5.8 [0–29]	<0.0001
UHDRS part IV, (/25 best) (<i>n</i> = 192) ^b	18.0 ± 7.1 [0–25]	22.6 ± 4.4 [5–25]	<0.0001
PHQ-9 (<i>n</i> = 164)			0.09
No depression ^a	95 (75%)	22 (59%)	
Mild or moderate depression ^a	16 (12.5%)	5 (14%)	
Major depression ^a	16 (12.5%)	10 (27%)	
EQ-5D VAS (/100 best) (<i>n</i> = 164) ^b	57% ± 21% [10–100]	63% ± 22% [20–100]	0.13
Wheelchair bound ^a	34 (24%)	6 (11%)	0.07

^aExpressed as number of patients (%). ^bExpressed as mean ± SD [range].

SARA = Scale for the assessment and rating of cerebellar ataxia; UHDRS = Unified Huntington Disease Rating Scale; PHQ-9 = Patient Health Questionnaire; EQ-5D VAS = Visual analog scale of the EQ-5D (self-rated health status).

limbs score was detailed only in patients who were able to walk (*n* = 172). Since controls scored 0 by definition, the discriminative value of this test is very low. Furthermore, the test is not specific for cerebellar dysfunction since the group with ADSP and ADCA did not differ [ADCA: 2.0 ± 2.2 (0–5), ADSP: 1.5 ± 1.9 (0–5), *p* NS].

Thirty-one patients were not able to perform all the upper limb tests: 26 could not write, 11 could not use the pegboard, seven could not perform the click test and four the tapping test. The patients who were able to perform all the tests had SARA scores <30. Among the seven most severely affected patients of our cohort with SARA values ≥30, none did perform all the tests: only five could perform the tapping test, four the nine-hole pegboard test (one only on the dominant side), four the click test (one only on the dominant side) and two the writing test. Patients with SARA values ≥30 were therefore excluded from subsequent analyses. Twenty-four additional patients did not perform one or more tests for different reasons: 12 tests were not done for reasons independent of the patient and four patients had unrelated physical difficulties (three were illiterate, hand deformity). These are completely at random failures that did not bias the results. There were eight patients out of 194 who were not able to do the tests because of their cerebellar syndrome or disease evolution (retinopathy and dystonic posture in two SCA7 patients) but had SARA scores ranging from 16 to 28. In order to test for relevance of this finding, we imputed the missing data and this did not modify the results.

The ADCA patients had significant higher *Z*-scores for each test compared to controls (*P*_c < 0.0001) and to the ADSP patients (*P*_c < 0.0001) (Table 3). There were no significant differences between controls and ADSP patients.

After linear regression analysis it seemed possible to identify the minimum number of items that would constitute a simple functional test capable of assessing disease severity as well as the SARA evaluation. This gave

rise to the composite cerebellar functional severity (CCFS) score, derived as follows.

CCFS score

In order to identify the minimal number of *Z*-scores that would account for the severity of the disease as measured by SARA in the ADCA patients, we performed a linear regression analysis of all of the functional tests. The results are shown in Table 4. In univariate analyses, all of the tests were significantly correlated with the SARA scores of the ADCA patients. In a stepwise multivariate linear regression analysis, only the pegboard and click tests on the dominant side were independently correlated with SARA scores. On these tests, each 1-point increase in SARA values was due to a 10- and 2.5-point increase of the respective *Z*-scores. Since the distribution of the CCFS score is not normal in the population of ADCA patients (data not shown), the scores were normalized by a decimal logarithmic transformation to allow discrimination of values in the lower ranges. Based on these results, we computed a composite score with the following formula

$$CCFS = \log_{10} \left(7 + \frac{Z_{\text{pegboard dominant hand}}}{10} + 4 \times \frac{Z_{\text{click dominant hand}}}{10} \right)$$

where 7 corresponds to the intercept of the multivariate linear regression equation (Table 4).

The results according to the genotype are given in Table 5. The mean CCFS score of all patients as a group was 1.06 ± 0.19 (0.73–1.6) with a median of 1.01. The CCFS was significantly higher (worse) in the ADCA group than in the ADSP group (1.12 ± 0.18 versus 0.90 ± 0.08, *P*_c < 0.0001), which in turn was higher (worse) than in the controls (0.85 ± 0.05, *P*_c < 0.03).

Table 3 Times and Z-scores on the four functional tests for controls or patients with ADSP or ADCA with SARA score below 30

Test	Times (s)		Z-scores		P-value ^a	
	All patients		Controls		All patients	
	All patients	ADCA	ADSP	ADCA	ADSP	ADCA
Tapping D	10.0 ± 5.8 [3;44] n = 186	0.1 ± 1.2 [-2.8;3.4] n = 123	3.9 ± 5.8 [-2.9;37.7] n = 187	0.7 ± 2.5 [-2.7;11.0] n = 50	5.1 ± 6.2 [-2.9;37.7] n = 137	<0.0001
Tapping non-D	10.5 ± 5.7 [5;38] n = 186	-0.1 ± 1.1 [-2.6;3.0] n = 123	4.4 ± 5.7 [-1.7;31.7] n = 187	1.4 ± 3.3 [-1.7;16.0] n = 50	5.5 ± 6.0 [-1.0;31.7] n = 137	<0.0001
Pegboard D	23.6 ± 19.4 [8;158] n = 181	0.0 ± 1.5 [-3.1;4.8] n = 123	12.9 ± 19.4 [-2.4;147.7] n = 182	2.1 ± 2.9 [-2.4;13.3] n = 50	17.0 ± 21.3 [-0.3;147.7] n = 132	<0.0001
Pegboard non-D	26.4 ± 19.0 [10;129] n = 181	-0.2 ± 1.6 [-3.6;6.5] n = 123	14.1 ± 19.0 [-2.2;117.2] n = 182	2.1 ± 3.0 [-2.2;14.5] n = 50	18.7 ± 20.4 [-0.5;117.2] n = 132	<0.0001
Click D	23.0 ± 16.4 [7;128] n = 185	0.1 ± 1.7 [-5.8;3.4] n = 123	12.7 ± 16.3 [-4.2;117.9] n = 186	2.1 ± 3.4 [-4.2;16.2] n = 50	16.5 ± 17.4 [-2.0;117.9] n = 136	<0.0001
Click non-D	24.4 ± 17.4 [5;128] n = 184	-0.1 ± 1.8 [-5.5;3.5] n = 123	13.3 ± 17.3 [-6.9;117.2] n = 185	1.9 ± 4.0 [-6.9;16.2] n = 50	17.5 ± 18.4 [-1.8;117.2] n = 135	<0.0001
Writing D	29.7 ± 25.7 [10;168] n = 166	0.0 ± 2.3 [-3.5;8.0] n = 123	18.8 ± 25.6 [-1.2;157.4] n = 167	5.2 ± 8.1 [-1.2;34.4] n = 47	24.1 ± 28.1 [-0.4;157.4] n = 120	<0.0001

All data are expressed in seconds as mean ± SD [Range]. D=dominant hand; nonD = non-dominant hand.

^aP-value for comparison between controls, ADSP and ADCA patients (ANOVA). Patients with ADCA differed significantly from controls and ADSP patients on all tests. ADSP patients did not differ from controls.

The CCFS was closely correlated, in patients with ADCA (Fig. 1) and ADSP (Fig. 2), with scores on the UHDRS-IV (ADCA: Pearson correlation coefficient $r = -0.78$, $P < 0.0001$ and ADSP: $r = -0.74$, $P < 0.0001$) and with the VAS of the EQ-5D (ADCA: $r = -0.41$, $P < 0.0001$ and ADSP: $r = -0.60$, $P < 0.0001$). It was significantly correlated with disease duration in ADCA ($r = 0.41$, $P < 0.0001$) but not in ADSP patients ($r = -0.002$, $P > 0.90$). In ADSP patients, CCFS was higher in patients with self-reported depression than in patients without (0.94 ± 0.11 versus 0.88 ± 0.05 , $P < 0.02$). In ADCA patients, the same tendency was observed but not statistically significant (1.16 ± 0.19 versus 1.10 ± 0.17 , $P > 0.07$).

When homogeneous genetic subgroups of ADCA and ADSP patients were analysed (Table 5), SCA3 patients had lower CCFS scores than SCA2 patients ($P_c < 0.04$). SPG4 patients had significantly lower scores than genetically undetermined ADSP patients (0.88 ± 0.05 versus 0.92 ± 0.09 , $P < 0.009$).

Discussion

This is the first study in which quantitative tests for cerebellar ataxia have been validated in controls. We showed in the controls that each functional test was significantly correlated with age. Consequently, all values were adjusted for age, and Z-scores, expressed as the difference between the observed time and the expected time due to age, were calculated to facilitate comparisons. This allowed us to study the rate of progression of the score independently of the age differences. This is an important methodological step in view of future clinical studies, since the inclusion of patients with different ages is unavoidable and changes in the score due to age must be distinguished from worsening due to the disease. Moreover, this will allow ageing to be taken into account in longitudinal studies.

The difficulty in our study was that the number of controls differed from the number of patients and the groups were in the same range but not exactly matched for age. The regression analysis we performed could therefore be biased, since it is more sensitive to the performance of the youngest and the oldest controls, which may mask ceiling effects. To avoid this, we examined linear, quadratic and cubic adjustments of the regression curves, with attention to the residuals. For the pegboard (for each hand separately), the quadratic model gave the best adjustment. As shown (Fig. 1, Supplementary data), this allowed us to take into account the ceiling effect for this test. For the other tests (tapping, click and writing), the best model was the linear model; the ceiling effect could be thus considered as small compared with increase of the performances with age. No clear increase of the residuals with age was shown.

Patients with ADCA performed significantly slower on all functional tests than patients with ADSP and controls, while ADSP patients and controls do not differ significantly. This showed that the tests were specific to cerebellar

Table 4 Correlations between functional test and SARA scores by univariate and multivariate linear regression analysis in patients with ADCA with SARA <30

	Univariate					Multivariate ($r^2 = 0.63$) $n = 115$		
	n	Parameter estimate	SE	P -value	r^2	Parameter estimate	SE	P -value
Intercept (multivariate)						6.7	0.61	
Tapping D	134	0.72	0.09	<0.0001	0.31			
Tapping non-D	134	0.86	0.09	<0.0001	0.41			
Pegboard D	129	0.32	0.03	<0.0001	0.50	0.12	0.04	<0.006
Pegboard non-D	130	0.26	0.03	<0.0001	0.43			
Click D	134	0.32	0.02	<0.0001	0.57	0.36	0.05	<0.0001
Click non-D	134	0.30	0.02	<0.0001	0.54			
Writing D	118	0.15	0.02	<0.0001	0.38			

Analyses were performed on the patients without missing data for the studied test (univariate analysis) or for all functional tests (multivariate analysis). SE = standard error of the parameter estimate.

Table 5 Comparison of the Composite Cerebellar Functional Severity Score (CCFS) in genetically homogeneous groups of patients with ADCA or ADSP

	Number of patients	CCFS Score (log seconds)	Age at onset (years)	Age at examination (years)	Disease Duration (years)
ADCA	141	1.12 ± 0.18 [0.81–1.64]	35.8 ± 12.3 [5–67]	47.3 ± 13.4 [15–83]	11.6 ± 7.7 [0–44]
SCA1	22	1.17 ± 0.26 [0.81–1.65]	40.3 ± 14.3 [17–67]	49.2 ± 15.5 [25–83]	9.2 ± 7.1 [0–23]
SCA2	32	1.19 ± 0.18 [0.93–1.59]	32.9 ± 12.3 [10–58]	45.6 ± 11.7 [19–69]	12.7 ± 9.2 [1–44]
SCA3	51	1.10 ± 0.16 ^a [0.86–1.50]	37.1 ± 10.2 [19–65]	48.5 ± 12.7 [23–74]	11.4 ± 6.3 [0–28]
ADSP	53	0.90 ± 0.08 [0.73–1.17]	22.8 ± 17.4 [0–56]	48.2 ± 14.2 [16–80]	24.9 ± 17.3 [2–69]
SPG4	22	0.88 ± 0.05 ^b [0.82–0.99]	27.3 ± 17.6 [0–56]	45.6 ± 13.9 [16–68]	17.4 ± 9.9 [2–35]
Controls	123	0.85 ± 0.05 [0.64–0.94]	–	45.1 ± 15.1 [20–74]	–

All data are expressed as mean ± SD [range].

^aSCA3 patients have significantly lower CCFS scores than SCA2 patients ($P_c < 0.04$). ^bSPG4 patients have significantly lower CCFS scores compared to other ADSP ($P < 0.009$).

involvement since in ADSP patients with the pure autosomal dominant form of the disease cerebellar signs are limited or absent. The small but non-significant difference could reflect the low mood in these patients.

Linear regression analysis showed that only two independent tests, the pegboard and the click tests on the dominant side, were sufficient to quantify the cerebellar signs of ADCA patients. We, therefore, developed a score incorporating the information from the two tests, the CCFS score, which clearly distinguishes cerebellar patients from patients with spastic paraplegias and controls. Importantly, the score is significantly correlated with disease duration but not subject to change in the presence of bad self-evaluated health status or low mood in ADCA. Therefore, the CCFS score is a measure of the disease, indirectly reflected by a poor quality of life and subjectively evaluated health, but not influenced by a depressed mood. Patients with most severe scores in click test are those who show the most severe hypermetric (overshoot) movements and the most delayed onset latencies of antagonist electromyographic activities during goal-directed movements (Manto *et al.*, 1994).

A potential limitation of the CCFS score is that severely affected patients might not be able to perform the tests.

This could be highly problematical in clinical trials where dropout of rates can be too high because the patients become too severely affected in the course of the trial. To avoid this problem, future trials should be designed for patients that can be evaluated throughout the course of the trial. Four out of the seven patients with SARA scores of 30 or more could not complete the pegboard and the click tests. This might be due to the length of time required to perform the task, which generates fatigue and distress. ADCA are very disabling disorders. In advanced stages of the disease and in addition to the cerebellar involvement, other systems are affected and examination reveals in the majority of cases pyramidal and extra-pyramidal signs. Interestingly, most severely affected patients were still able to perform the tapping test. Moreover, since they performed similarly with both hands on this test, it can be used in more severely affected patients to assess their evolution. It will be necessary to determine whether this test remains discriminative in severely affected patients. In addition, the finding that patients with SARA scores >30 are not able to perform the tests for our functional composite score suggests that this might be a useful end-point for future studies.

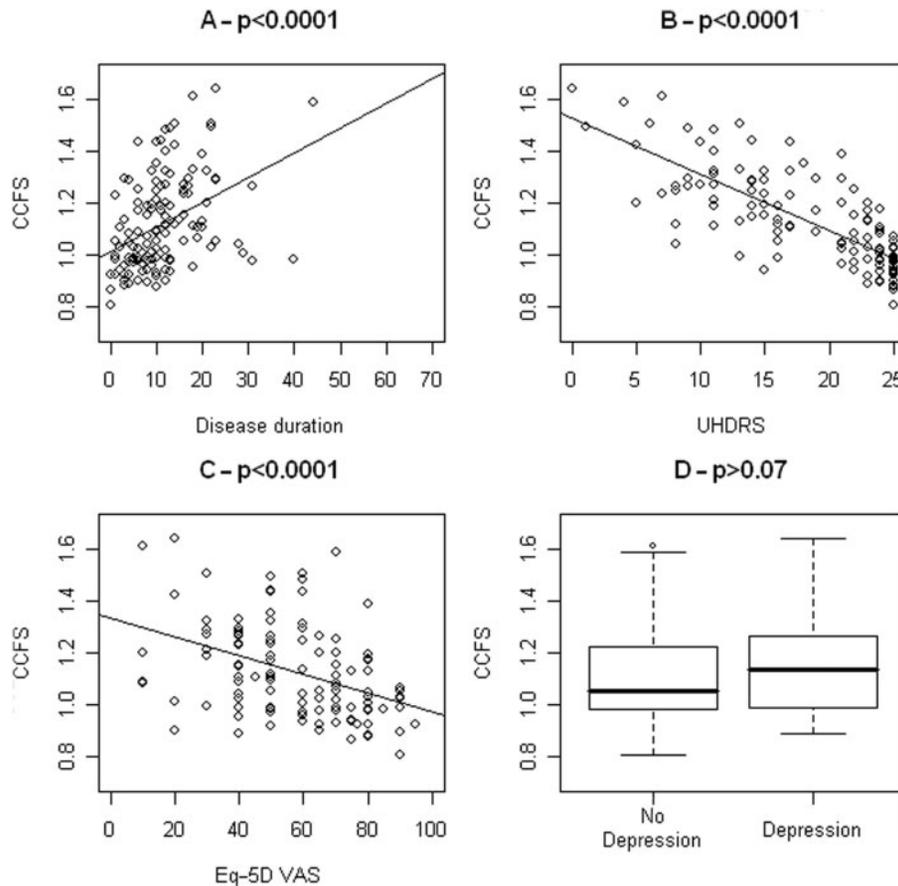


Fig. 1 Linear regression plots showing correlations in ADCA patients between the CCFS scores and other disease-related parameters. UHDRS = part IV of the Unified Huntington disease Rating Scale; EQ-5D VAS = Visual analogue scale of the EQ-5D (self-rated health status).

Missing values could introduce a bias in the interpretation of our results. Missing values related to the disease (i.e. visual problems caused by the disease, too severe disease) were imputed without changing the results of the statistical analysis. Nevertheless, this could introduce a bias in a large-scale trial, especially with very disabled patients. The discovery of biomarkers that detect biochemical changes rather than motor impairment is an important challenge for future investigations.

Patients with ADSP could be distinguished from controls by their CCFS scores and not surprisingly, even more by their SARA scores, which assess more than just cerebellar features. Nineteen per cent of the ADSP patients had SARA scores >7.5 , the highest value seen in a control population (5.8 ± 5.8 versus 0.4 ± 1.1 according to Schmitz-Hübsch *et al.*, 2006b). This might mean that in ADSP patients, a cerebellar dysfunction affects the upper limbs. However, general slowness due to other factors might also explain the difference. Depression is the most evident such factor, since it was more frequent in ADSP patients than in those with ADCA (Table 1), although their functional impairment (wheelchair use or UHDRS scores, Table 1) was less severe. This observation is important because it demonstrates that

even a very specific ataxia score, such as SARA, can be influenced by more global impairment.

The high frequency of depression among ADSP patients shows that they need psychiatric care and psychological help as much as cerebellar patients, who are objectively more physically impaired. CCFS was correlated with self-evaluated depressive mood in ADSP patient only.

Among genetically distinct forms of cerebellar ataxia, CCFS scores were lower in SCA3 patients than in SCA1 and SCA2 patients (Table 5). These results suggest that SCA3 might progress more slowly than SCA1 and SCA2; calculation of the CCFS in a prospective follow-up study of cerebellar ataxias will allow us to confirm this hypothesis.

The CCSF has several advantages over other evaluation procedures for cerebellar ataxia. (i) It is easy to perform; (ii) it is based on timed activities, the absence of which is a drawback for prospective studies, according to a criticism of the SARA evaluation (Subramony, 2007); (iii) only upper limb function is quantified, but as the disease progresses patients lower limbs become too poorly coordinated to perform a test; and (iv) most importantly, pyramidal signs and peripheral neuropathy (Stevanin *et al.*, 2002; Kubis *et al.*, 1999; van de Warrenburg *et al.*, 2004) are frequent in

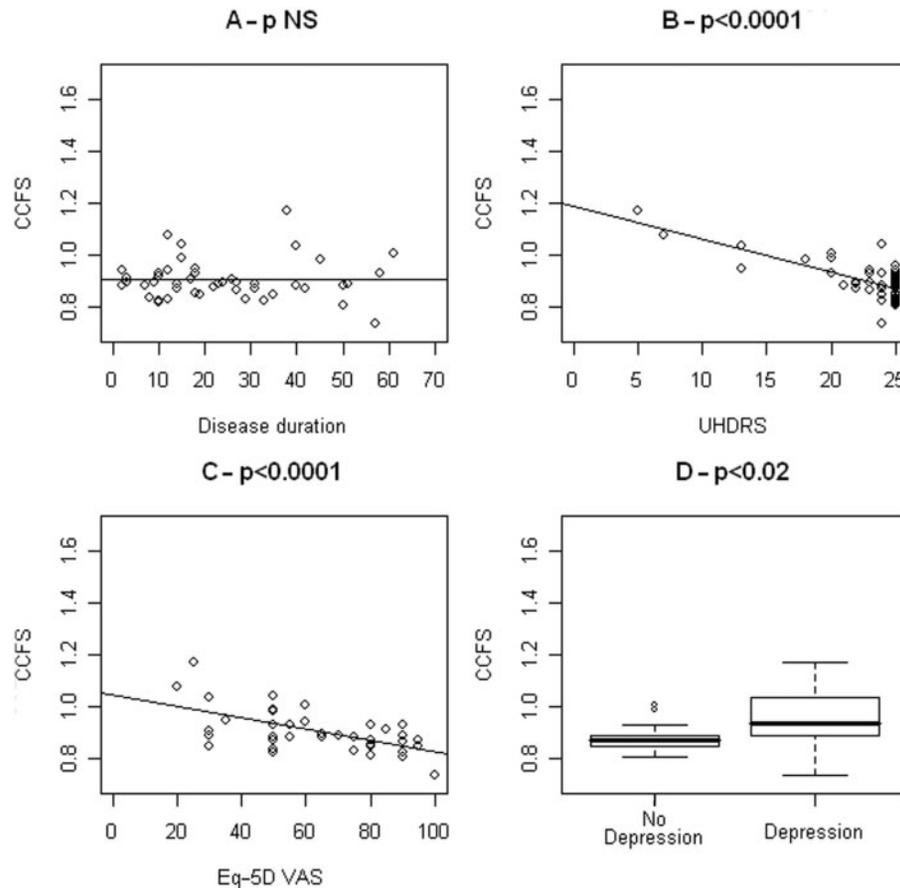


Fig. 2 Linear regression plots showing correlations in ADSP patients between the CCFS scores and other disease-related parameters. UHDRS = part IV of the Unified Huntington disease Rating Scale; EQ-5D VAS = Visual analogue scale of the EQ-5D (self-rated health status); NS = not significant.

patients with cerebellar ataxia and may interfere with the evaluation of cerebellar signs in the lower limbs.

There are limitations to the CCFS. Testing only the performance of upper limbs might be thought not to adequately reflect the disease. The cerebellar syndrome in degenerative ataxias begins in the axis and affects the ability to stand, followed by the kinetic disorders in legs and arms. Since the slopes of the correlations with disease duration of the SARA standing score (item 1) and the SARA upper limb score (item 6) showed that the former evolved two times faster than the latter (Fig. 2, Supplementary data), we concluded that an upper limb test could be used longer in ADCA patients than a lower limb test. Thus, we would be able to include more patients in a potential trial.

Although incorporation of a lower limb score in the composite score could be potentially very useful, when we included a lower limb evaluation, we lost a substantial number of patients; 24% of the ADCA and 11% of ADSP were wheelchair bound and could not perform the walking test. In addition, the CCFS was correlated with the lower limb score, showing that upper and lower limb performances are not completely independent of each other.

Nevertheless, other tests could be used to assess progressive symptoms such as dysarthria for example. The PATA test assesses speech speed (repetitions of 'PATA' in a 10-s interval using a tape recorder to examine speech). In a study on Friedreich ataxia, the authors showed a good correlation for the pegboard and the PATA score and the severity of the disease (Subramony *et al.*, 2005). On the contrary, the PATA score showed that some raters scored consistently higher or lower than others. In addition, speech speed is valuable, but multilingual patient groups could be problematic in an international study designs with multi-centre settings.

As previously shown by our group, oculomotor recordings showed typical patterns and evolutions in SCA subgroups (Rivaud-Pechoux *et al.*, 1998). This could be used to assess changes over time during trials. Since changes are specific of each SCA the recording should be done in genetically defined subgroups. Moreover, a minimal equipment is necessary to make the recordings, which could limit its use in some trial situations.

One of the advantages of other composite measures is that they capture neurological dysfunction regardless of the anatomical abnormality; in other words, they represent

measures of disease progression not cerebellar dysfunction. Thus, CCFS could be a useful or a problematic score, depending on the population assessed and the nature of the trial. The cerebellar dysfunctions in the upper limbs could miss more widespread neurological dysfunction. A treatment that improves cerebellar dysfunction, as measured by our score, might not improve other signs that are experienced as more handicapping by the patients. An evaluation of changes in the CCFS changes in a longitudinal study will allow us to test for sensitivity to changes over time with regard to the cerebellar impairment versus overall disease progression.

Supplementary data

Supplementary material is available at *Brain* online.

Acknowledgements

We warm fully thank all the patients and controls for their participation in the study. We thank also Lydia Guennec for entering the data for statistical analysis. This study was funded by PHRC AOM 0359 (AP-HP), INSERM ANR (contract n°A05221DS) and EuroSca (contract n°LSHM-CT-2004-503304).

References

- Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* 1999; 122: 871–82.
- David G, Abbas N, Stevanin G, Dürr A, Yvert G, Cancel G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* 1997; 17: 65–70.
- Depienne C, Stevanin G, Brice A, Dürr A. Spastic paraplegias: an update. *Curr Opin Neurol* 2007; 20: 674–80.
- EuroQoL Group. EuroQoL: a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199–208.
- Harding AE. Clinical features and classification of inherited ataxias. *Adv Neurol* 1993; 61: 1–14.
- Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 1996; 11: 136–42.
- Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, et al. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG glutamine repeats. *Nat Genet* 1996; 14: 285–91.
- Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, et al. CAG expansions in a novel for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 1994; 8: 221–8.
- Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, et al. Unstable expansion of CAG repeat in hereditary dentato-rubral-pallidolusian atrophy (DRPLA). *Nat Genet* 1994; 6: 9–13.
- Kubis N, Dürr A, Gugenheim M, Chneiweiss H, Mazzetti P, Brice A, et al. Polyneuropathy in autosomal dominant cerebellar ataxias: phenotype-genotype correlation. *Muscle Nerve* 1999; 22: 712–7.
- Lynch DR, Farmer JM, Tsou AY, Perlman S, Subramony SH, Gomez CM, et al. Measuring Friedreich ataxia: complementary features of examination and performance measures. *Neurology* 2006; 66: 1711–6.
- Manto M, Godaux E, Jacqy J. Cerebellar hypermetria is larger when the inertial load is artificially increased. *Ann Neurol* 1994; 35: 45–52.
- Mathiowetz V, Weber K, Kashman B, Volland G. Adult norms for the nine hole peg test of finger dexterity. *Am J Occup Ther* 1985; 5: 24–38.
- Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, et al. Dentatorubral and pallidolusian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat Genet* 1994; 6: 14–8.
- Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, et al. Expansion of an unstable trinucleotide CAG repeats in spinocerebellar ataxia type 1. *Nat Genet* 1993; 4: 221–6.
- Rivaud-Pechoux S, Dürr A, Gaymard B, Cancel G, Ploner CJ, Agid Y, et al. Eye movement abnormalities correlate with genotype in autosomal dominant cerebellar ataxia type I. *Ann Neurol* 1998; 43: 297–302.
- Sanpei K, Takano F, Igarashi S, Sato T, Oyake M, Sasaki H, et al. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 1996; 14: 277–84.
- Schmitz-Hübsch T, Tezenas du Montcel S, Baliko L, Boesch S, Bonato S, Fancellu R, et al. Reliability and validity of the International Cooperative Ataxia Rating Scale: a study in 156 spinocerebellar ataxia patients. *Mov Disord* 2006a; 21: 699–704.
- Schmitz-Hübsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* 2006b; 66: 1717–20. Erratum in: *Neurology* 2006; 67: 299.
- Schols L, Amoiridis G, Buttner T, Przuritek H, Epplen JT, Riess O. Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes? *Ann Neurol* 1997; 42: 924–32.
- Spitzer R, Kroenke K, Williams J. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. *JAMA* 1999; 282: 1737–44.
- Stevanin G, Dürr A, Brice A. Spinocerebellar ataxias caused by polyglutamine expansions. *Adv Exp Med Biol* 2002; 516: 47–77.
- Subramony S. SARA- a new clinical scale for the assessment and rating of ataxia. *Nature clinical practice Neurology* 2007; 3: 126–37.
- Subramony SH, May W, Lynch D, Gomez C, Fischbeck K, Hallett M, et al. Cooperative Ataxia Group. Measuring Friedreich ataxia: interrater reliability of a neurologic rating scale. *Neurology* 2005; 64: 1261–2.
- Trouillas P, Takayanagi T, Hallett M, Currier RD, Subramony SH, Wessel K, et al. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropharmacology Committee of the World Federation of Neurology. *J Neurol Sci* 1997; 145: 205–11.
- van de Warrenburg BP, Notermans NC, Schelhaas HJ, van Alfen N, Sinke RJ, Knoers NV, et al. Peripheral nerve involvement in spinocerebellar ataxias. *Arch Neurol* 2004; 61: 257–61.
- Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, et al. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the 1A-voltage-dependent calcium channel. *Nat Genet* 1997; 15: 62–9.