



Combined cerebral atrophy score in Huntington's disease based on atlas-based MRI volumetry: Sample size calculations for clinical trials

Hans-Peter Müller^{a,*,1}, Hans-Jürgen Huppertz^{b,1}, Jens Dreyhaupt^c, Albert C. Ludolph^a, Sarah J. Tabrizi^d, Raymund A.C. Roos^e, Alexandra Durr^f, G. Bernhard Landwehrmeyer^{a,2}, Jan Kassubek^{a,2}

^a Department of Neurology, University of Ulm, Germany

^b Swiss Epilepsy Clinic, Klinik Lengg, Zürich, Switzerland

^c Institute of Epidemiology and Medical Biometry, University of Ulm, Germany

^d Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK

^e Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands

^f ICM - Institut du Cerveau et de la Moelle Epinière, INSERM U1127, CNRS UMR7225, Sorbonne Universités – UPMC Université Paris VI UMR_S1127 and APHP, Genetic Department, Pitié-Salpêtrière University Hospital, Paris, France

ARTICLE INFO

Keywords:

Atlas-based volumetry
Clinical trial
Effect size
Huntington's disease
Longitudinal study
Multicenter study
Surrogate marker

ABSTRACT

Introduction: A volumetric MRI analysis of longitudinal regional cerebral atrophy in Huntington's disease (HD) was performed as a read-out of disease progression to calculate sample sizes for future clinical trials.

Methods: This study was based on MRI data of 59 patients with HD and 40 controls recruited within the framework of the PADDINGTON study and investigated at baseline and follow-up after 6 and 15 months. Automatic atlas-based volumetry (ABV) of structural T1-weighted scans was used to calculate longitudinal volume changes of brain structures relevant in HD and to assess standardized effect sizes and sample sizes required for potential future studies.

Results: Atrophy rates were largest in the caudate (−3.4%), putamen (−2.8%), nucleus accumbens (−1.6%), and the parietal lobes (−1.7%); the lateral ventricles showed an expansion by 6.0%. Corresponding effect sizes were −1.35 (caudate), −0.84 (putamen), −0.91 (nucleus accumbens), −1.05 (parietal lobe), and 0.92 (lateral ventricles) leading to $N = 36$ subjects per study group for detecting a 50% attenuation of atrophy for the best performing structure (caudate). A combined score of volume changes in non-overlapping compartments (striatum, parietal lobes, lateral ventricles) increased the effect size to −1.60 and substantially reduced the required sample sizes by 10 to $N = 26$ subjects per study group. This combined imaging score correlated significantly both with the CAP score and with the progression of the clinical phenotype.

Conclusion: We propose ABV of the striatum together with parietal lobe and lateral ventricle volumes as a combined imaging read-out for progression studies including clinical trials in HD.

1. Introduction

Huntington's disease (HD) is a progressive neurodegenerative disease for which currently only symptomatic treatment is available [1]. Future clinical trials aiming at disease modification in HD need sensitive *in vivo* biomarkers to track disease progression [2]. Structural neuroimaging, as from PREDICT [3], TRACK-HD [4,5], and PADDINGTON [6], demonstrated striatal atrophy starting as early as 15 years before clinical onset and continuing throughout the pre-manifest

periods. The extent of striatal atrophy adds predictive power for motor onset beyond age and CAG repeat length alone [3]. HD-associated basal ganglia atrophy [7], additional atrophy in the deep gray matter including the thalamus [8], hippocampus, nucleus accumbens, and amygdala [9], or atrophy of the corpus callosum [6] were observed *in vivo* from magnetic resonance imaging (MRI) data. The potential of MRI-based findings in HD in treatment trials has already been summarized, suggesting caudate volume as a potential biomarker [2].

In an approach unbiased by a *priori* assumptions, we aimed to

* Corresponding author. Department of Neurology, University of Ulm, Oberer Eselsberg 45, D-89081, Ulm, Germany.

E-mail address: hans-peter.mueller@uni-ulm.de (H.-P. Müller).

¹ Shared first authorship.

² Shared senior authorship.

analyze if there are volumetric parameters which would perform even better as outcome measure, thus allowing trials with fewer patients despite comparable power to detect therapeutic efficacy. To this end, the fully automatic approach of atlas based volumetry (ABV) [10–12] was used to quantify longitudinal regional brain volume changes. ABV is an objective, investigator-independent technique with low intrascanner variability [10,13] to determine the volumes of intracranial compartments and cerebral substructures from the MRI data of individual subjects. The method was applied in order to calculate HD-associated annualized percentage cerebral volume changes, corresponding standardized effect sizes, and to estimate sample sizes required for potential future disease-modifying trials, targeting the definition of an imaging based combined progression score.

2. Methods

2.1. Data recording and subjects

Sixty-one patients with HD and 40 controls were enrolled into the prospective, longitudinal, cohort observational study *Pharmacodynamic approaches to demonstration of disease-modification in Huntington's disease* (PADDINGTON) [14] at Leiden (the Netherlands), London (UK), Paris (France), and Ulm (Germany). Assessments and MRI acquisitions were performed at baseline, 6 and 15 months.

Patients were recruited from research centers; controls were spouses, partners or gene-negative siblings in order to match patients to controls as closely as possible in terms of age, education level, background, and home life. All participants were ambulatory and agreed to volunteer for MRI scanning and ensuing data analysis. The study was approved by the local ethical committees of the sites involved.

MRI scans and further clinical data were acquired at the same visit. HD patients had a genetically confirmed disease with a trinucleotide (CAG) repeat length of 36 or higher, and had clinical features of mild HD at stage I based on the Unified Huntington's disease Rating Scale (UHDRS) [15] with a TFC value of 11–13, indicating good capacity in functional realms [14].

T1-weighted scans were acquired at 3 Tesla with sagittal slices of 1.1 mm thickness, with no inter-slice gaps, and an in-plane resolution of approximately $1.1 \times 1.1 \text{ mm}^2$. MRI acquisitions were performed longitudinally at baseline, after 6 months, and after 15 months, as previously described [14]. Subjects' distribution and scan parameters are listed in Table 1. Thorough quality control of each scan was performed, and scans with obvious factors biasing volumetry such as movement artefacts were excluded. From the 61 HD patients and the 40 controls initially included in this study, 59 HD patients and 40 controls contributed to the results. In detail, 13 data sets from HD patients and three data sets from controls did not contribute due to missing data or data that did not pass the quality control. If only two data sets of the intended three data sets were available, volume changes were calculated from these two data sets – this was the case for nine HD patients (out of 59) and three controls (out of 40). Two HD patients were completely

excluded because less than two MRIs were available. Thus, although based on the same acquired data set, the results of this study could not directly be compared to the results of a former MRI analysis by Hobbs and coworkers [6].

2.2. Atlas-based volumetry

All MPRAGE data were processed by use of MATLAB (R2014b, The Mathworks, USA) using the Statistical Parametric Mapping 12 (SPM12) software (Wellcome Trust Center for Neuroimaging, London, UK, www.fil.ion.ucl.ac.uk/spm) according to a standardized processing pipeline for ABV [12]. Briefly, processing includes (i) segmentation into gray matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) compartments, (ii) stereotaxic normalization into Montreal Neurological Institute (MNI) space, (iii) ABV using voxel-by-voxel multiplication and subsequent integration of normalized modulated component images (GM, WM or CSF) with predefined masks from different brain atlases. To enhance the quality of mapping into atlas space, high-dimensional registration methods have recently been introduced, and the intrascanner variability of volumetric results was shown to be $< 1\%$ for the majority of investigated structures [13]. The method of ABV has been successfully employed in cross-sectional and longitudinal studies [11,12,16,17]. Specifically in the current study, the T1-weighted scans of each subject have been coregistered by means of the Longitudinal Registration Toolbox of SPM12 prior to normalization, segmentation and subsequent volumetric analysis. This has the advantage that the stereotactic normalization to MNI space has to be done only once, i.e., for the so-called mid-point average image resulting from longitudinal registration, which reduces measurement variability [18].

For the purpose of the current study, the volumes of 38 structures and compartments were analyzed (cf. Table 2). The masks for ABV were derived from different probabilistic brain atlases: the Harvard-Oxford atlas of subcortical structures distributed with the Oxford Center for Functional MRI of the Brain Software Library (FSL) package [19] for hippocampus, caudate, putamen, nucleus accumbens, pallidum, and thalamus; the Hammersmith atlas n30r83 [20] for the third and lateral ventricles; the LONI Probabilistic Brain Atlas [21] for all other structures. The volume of the striatum was summed up from caudate nucleus, putamen and nucleus accumbens. GM, WM, and CSF volumes and intracranial volume (ICV) were determined by the “tissue volumes” utility of SPM12 [22]. All individual cross-sectional structural volumes as assessed by ABV were normalized for individual head size for each subject as follows: first, the individual result volumes were divided by the individual ICV (expressing each structure volume as percentage of ICV) and second, these percentage results were then multiplied by the mean ICV value of all subjects (which was 1400 ml):

$$V_{norm} = V_{analysis} / V_{ICV} * 1400 \text{ ml} \quad (1)$$

where V_{norm} denotes the normalized volume of a subject's brain structure $V_{analysis}$ the measured volume of a subject's brain structure, and V_{ICV} the intracranial volume (all measures in ml), while the target

Table 1

Distribution of HD patients and controls (age- and gender-matched) from the different sites contributing to the results of this study, with age ranges (mean \pm standard deviation (minimum - maximum)), centerwise ratios between HD patients/controls, MR scanners and MR sequences. All subjects were investigated at the same field strength of 3.0 T with similar spatial resolution, different echo time (TE), and repetition time (TR) resulting in equalized image contrasts.

Distribution of HD patients and controls from the different sites										
Center	HD patients		Controls		Ratio HD patients/controls	MR scanner	MR sequence matrix	resolution/mm ³	TE/ms	TR/ms
	m/f	age/years	m/f	age/years						
A	2/15	48 \pm 10 (32–67)	6/4	49 \pm 6 (38–57)	17/10	Philips Achieva 3.0 T	166 \times 240 \times 240	1.1 \times 1.0 \times 1.0	3.5	7700
B	4/12	52 \pm 9 (40–67)	4/6	53 \pm 7 (44–66)	16/10	Siemens Tim Trio 3.0 T	207 \times 256 \times 256	1.1 \times 1.1 \times 1.1	2.9	2200
C	5/6	45 \pm 15 (23–63)	3/7	57 \pm 8 (45–66)	11/10	Siemens Verio 3.0 T	207 \times 256 \times 256	1.1 \times 1.1 \times 1.1	2.9	2200
D	11/4	48 \pm 10 (30–64)	4/6	46 \pm 9 (29–59)	15/10	Siemens Allegra 3.0 T	207 \times 256 \times 256	1.1 \times 1.1 \times 1.1	3.7	2200
All	22/37	49 \pm 11 (23–67)	17/23	51 \pm 8 (29–66)	59/40					

Table 2

Cross-sectional volumetric results, longitudinal volume changes, effect sizes, and sample sizes in HD patients and controls. The cross-sectional results have been normalized to a mean intracranial volume (ICV) of 1400 ml, the longitudinal results have not been ICV corrected. The relative difference HD to controls was calculated by the difference of group averaged volumes divided by the mean volume of controls. A 3-color scale was used to rank the relative volume differences to controls from shades of red (volume loss) over white to shades of blue (volume gain). The same 3-color scale was used to rank the standardized effect sizes from shades of red (negative effect size) over white to shades of blue (positive effect size). A Student's t-test was applied to investigate the significance of annualized volume changes between HD patients and controls, with Bonferroni-Holm correction of p values for multiple comparisons. Red colored p values mark significance. The treatment effect columns display the required number of patients in each study arm (verum or placebo) for expected treatment effects of 20%, 30% or 50%, respectively. A 2-color scale was used to rank the results in these 3 columns from higher numbers (white) to lower numbers (shades of green). SD – standard deviation; GM – gray matter; WM – white matter; CI – confidence interval.

structures / compartments	cross-sectional results					longitudinal results														
	controls		HD patients		relative difference HD to controls	controls		HD patients		standardized effect sizes	p (annualized volume changes HD vs controls)	treatment effect								
	baseline ICV corrected		baseline ICV corrected			relative annualized volume change (95%CI)		relative annualized volume change (95%CI)				20%	30%	50%						
	mean (ml)	SD (ml)	mean (ml)	SD (ml)		mean	SD	mean	SD											
global																				
brain	1096.9	38.0	1038.8	64.3	-5.30%	0.20% (-0.24%;0.64%)	1.73%	-1.07% (-1.45%;-0.68%)	1.50%	-0.80	0.000	>200	>200	101						
gray matter (GM)	663.0	39.0	626.3	60.0	-5.53%	-0.06% (-0.56%;0.43%)	1.93%	-1.38% (-1.91%;-0.84%)	2.09%	-0.65	0.002	>200	>200	151						
white matter (WM)	433.9	26.5	412.5	26.7	-4.94%	0.61% (0.06%;1.15%)	2.15%	-0.59% (-1.11%;-0.08%)	2.02%	-0.58	0.007	>200	>200	189						
CSF	303.1	38.0	361.2	64.3	19.16%	0.88% (-0.26%;2.02%)	4.45%	3.31% (2.16%;4.45%)	4.50%	0.54	0.010	>200	>200	>200						
intracranial volume	1377.2	118.2	1352.3	141.1	-1.81%	0.32% (-0.10%;0.74%)	1.64%	0.05% (-0.25%;0.35%)	1.17%	0.51	0.013	>200	>200	>200						
cerebral lobes																				
cerebrum	941.8	36.5	885.2	60.5	-6.01%	0.02% (-0.37%;0.41%)	1.54%	-1.29% (-1.66%;-0.93%)	1.42%	-0.89	0.000	>200	>200	80						
cerebrum GM	556.5	35.2	519.9	55.1	-6.57%	-0.29% (-0.71%;0.14%)	1.67%	-1.71% (-2.26%;-1.17%)	2.15%	-0.72	0.000	>200	>200	121						
cerebrum WM	385.4	25.0	365.3	25.2	-5.20%	0.46% (-0.08%;1.01%)	2.12%	-0.69% (-1.21%;-0.17%)	2.04%	-0.56	0.009	>200	>200	>200						
frontal lobe	309.4	19.3	293.6	23.6	-5.13%	-0.11% (-0.47%;0.25%)	1.42%	-1.31% (-1.79%;-0.83%)	1.88%	-0.70	0.000	>200	>200	129						
frontal lobe GM	180.9	15.4	170.6	19.6	-5.69%	-0.42% (-0.85%;0.01%)	1.68%	-1.71% (-2.40%;-1.02%)	2.70%	-0.55	0.004	>200	>200	>200						
frontal lobe WM	128.5	10.1	122.9	10.4	-4.34%	0.33% (-0.19%;0.86%)	2.05%	-0.71% (-1.32%;-0.09%)	2.40%	-0.46	0.023	>200	>200	>200						
temporal lobe	189.1	9.2	179.6	13.7	-5.03%	0.37% (-0.39%;1.13%)	2.95%	-0.83% (-1.43%;-0.23%)	2.34%	-0.46	0.036	>200	>200	>200						
temporal lobe GM	134.4	8.7	127.5	12.6	-5.13%	0.24% (-0.53%;1.00%)	3.01%	-1.01% (-1.69%;-0.34%)	2.64%	-0.45	0.036	>200	>200	>200						
temporal lobe WM	54.7	4.0	52.1	3.9	-4.80%	0.68% (-0.22%;1.57%)	3.50%	-0.39% (-1.15%;0.38%)	2.99%	-0.33	0.119	>200	>200	>200						
parietal lobe	173.4	9.6	164.6	14.0	-5.03%	0.05% (-0.37%;0.46%)	1.62%	-1.68% (-2.10%;-1.25%)	1.66%	-1.05	0.000	>200	161	59						
parietal lobe GM	103.3	7.4	97.1	12.3	-5.95%	-0.44% (-0.87%;-0.01%)	1.68%	-2.25% (-2.90%;-1.60%)	2.54%	-0.81	0.000	>200	>200	97						
parietal lobe WM	70.1	5.7	67.5	6.2	-3.69%	0.76% (0.09%;1.44%)	2.64%	-0.83% (-1.51%;-0.15%)	2.66%	-0.60	0.004	>200	>200	176						
occipital lobe	125.0	8.4	115.6	10.9	-7.54%	-0.26% (-0.94%;0.42%)	2.66%	-1.77% (-2.35%;-1.19%)	2.28%	-0.62	0.004	>200	>200	165						
occipital lobe GM	79.5	6.9	72.6	9.6	-8.58%	-0.56% (-1.34%;0.21%)	3.04%	-2.15% (-2.91%;-1.38%)	3.00%	-0.53	0.012	>200	>200	>200						
occipital lobe WM	45.6	4.4	43.0	4.0	-5.71%	0.30% (-0.50%;1.11%)	3.15%	-1.11% (-1.89%;-0.33%)	3.05%	-0.46	0.029	>200	>200	>200						
subcortical structures																				
insula	17.3	1.3	16.5	1.6	-4.68%	-0.10% (-0.56%;0.36%)	1.80%	-1.42% (-1.90%;-0.93%)	1.90%	-0.71	0.001	>200	>200	128						
cerebellum	115.5	10.0	115.3	9.6	-0.20%	1.38% (0.17%;2.59%)	4.74%	0.47% (-0.46%;1.40%)	3.66%	-0.22	0.307	>200	>200	>200						
cerebellum GM	93.7	9.1	93.7	8.4	0.06%	1.20% (-0.07%;2.46%)	4.96%	0.32% (-0.61%;1.25%)	3.64%	-0.21	0.340	>200	>200	>200						
cerebellum WM	21.8	1.7	21.6	2.4	-1.32%	2.14% (0.98%;3.30%)	4.55%	1.05% (-0.24%;2.34%)	5.05%	-0.22	0.266	>200	>200	>200						
brainstem	29.7	2.1	28.8	2.5	-3.27%	1.28% (0.36%;2.19%)	3.57%	-0.01% (-0.66%;0.64%)	2.56%	-0.43	0.054	>200	>200	>200						
midbrain	10.3	0.6	9.9	0.6	-3.73%	0.72% (0.11%;1.33%)	2.40%	-0.38% (-0.86%;0.10%)	1.88%	-0.52	0.018	>200	>200	>200						
pons	14.9	1.3	14.4	1.7	-3.52%	1.47% (0.45%;2.50%)	4.01%	0.12% (-0.62%;0.86%)	2.89%	-0.40	0.071	>200	>200	>200						
medulla	4.5	0.4	4.5	0.4	-1.51%	1.86% (0.41%;3.32%)	5.70%	0.39% (-0.55%;1.32%)	3.66%	-0.32	0.153	>200	>200	>200						
hippocampus	6.8	0.5	6.5	0.6	-5.24%	0.61% (-0.03%;1.25%)	2.52%	-0.32% (-0.77%;0.14%)	1.78%	-0.44	0.048	>200	>200	>200						
amygdala	3.8	0.3	3.5	0.3	-7.73%	0.67% (-0.03%;1.37%)	2.75%	-0.41% (-0.92%;0.11%)	2.03%	-0.46	0.038	>200	>200	>200						
striatum	11.7	1.2	8.3	1.6	-29.67%	0.24% (-0.46%;0.94%)	2.73%	-2.95% (-3.66%;-2.24%)	2.79%	-1.15	0.000	>200	133	49						
caudate	4.6	0.4	3.1	0.7	-32.62%	0.02% (-0.41%;0.45%)	1.68%	-3.38% (-4.13%;-2.63%)	2.95%	-1.35	0.000	>200	97	36						
putamen	6.1	0.8	4.3	0.9	-28.70%	0.46% (-0.60%;1.51%)	4.14%	-2.84% (-3.80%;-1.89%)	3.76%	-0.84	0.000	>200	>200	90						
accumbens	1.0	0.1	0.8	0.1	-22.20%	0.16% (-0.43%;0.75%)	2.31%	-1.65% (-2.09%;-1.20%)	1.74%	-0.91	0.000	>200	>200	78						
pallidum	3.8	0.3	3.3	0.2	-14.32%	0.34% (-0.13%;0.80%)	1.82%	-1.10% (-1.45%;-0.75%)	1.37%	-0.92	0.000	>200	>200	76						
thalamus	12.7	0.8	12.0	1.1	-5.70%	0.33% (-0.13%;0.79%)	1.80%	-0.80% (-1.18%;-0.41%)	1.50%	-0.69	0.002	>200	>200	132						
ventricles																				
lateral ventricles	14.3	6.8	19.2	8.0	34.46%	2.29% (1.28%;3.30%)	3.95%	6.01% (4.97%;7.06%)	4.09%	0.92	0.000	>200	>200	75						
third ventricle	0.8	0.3	1.2	0.4	47.01%	2.19% (1.08%;3.30%)	4.36%	4.48% (3.47%;5.50%)	3.97%	0.56	0.009	>200	>200	>200						
combination score (parietal lobe, striatum, lateral ventricles)																				
C	63.8	22.2	34.9	25.0	-45.21%	-1.79% (-2.91%;-0.68%)	4.37%	-9.70% (-11.05%;-8.35%)	5.28%	-1.60	0.000	155	69	26						

volume of 1400 ml approximates the average ICV of adult subjects. All longitudinal evaluations in this study, however, were based on absolute volumetric results, i.e., without ICV correction. By calculating volume differences between different points in time, possible effects of head size variations are already sufficiently eliminated. Moreover, the calculation of the ICV involves the CSF volume whose separation from bone and determination in T1-weighted images is not easy [22]. This could have introduced a measurement inaccuracy that might be larger than the small volume differences that we wanted to determine in the longitudinal evaluations.

2.3. Calculation of volume changes

Volume changes for each structure or compartment were calculated by linear regression analysis of the volumes at the three visits. Relative annualized volume changes $P_{\Delta V}$ were expressed as volume changes relative to the initial volume (at baseline) and normalized to one year:

$$P_{\Delta V} = [(V_2 - V_1)/(t_2 - t_1) + (V_3 - V_1)/(t_3 - t_1) + (V_3 - V_2)/(t_3 - t_2)] / (3V_1) \quad (2)$$

where V_i denotes the volume at the respective visit at time t_i (in years with two decimal places).

2.4. Calculation of standardized longitudinal effect sizes

Standardized effect sizes were calculated by the differences of mean volume change in HD patients and in controls, divided by the weighted mean standard deviation of volume changes in HD patients and controls. Beyond the single structures, combinations of measurements for the different structures as combined imaging read-out were analyzed in order to increase the effect sizes and to achieve the least number of patients for detecting efficacy to slow down disease progression. We selected the following non-overlapping structures/compartments with high individual effect sizes for a combined score: the striatum (as a subcortical structure with relevant alteration in HD), the parietal lobes (as a marker of cortical/lobar involvement in HD), and the lateral ventricles (as an indirect marker of cerebral atrophy by *e vacuo* expansion). For each individual subject, the volumes of striatum and parietal lobes were multiplied and divided by the volume of the lateral ventricles since the enlargement of the CSF space runs counter the atrophy of the brain parenchyma. That way, a data-driven approach was chosen to include gray matter and white matter structures (striatum), the ventricle system and a global lobe structure (parietal lobe):

$$C = V(\text{parietal lobe}) * V(\text{striatum}) / V(\text{lateral ventricle}). \quad (3)$$

For the combined parameter C, Eq. (2) alters to

$$P_{\Delta C} = [(C_2 - C_1)/(t_2 - t_1) + (C_3 - C_1)/(t_3 - t_1) + (C_3 - C_2)/(t_3 - t_2)] / (3C_1). \quad (4)$$

A Student's *t*-test was applied to investigate the significance of annualized volume changes between HD patients and controls, with Bonferroni-Holm correction of *p* values for multiple comparisons.

2.5. Sample size calculations

The sample size calculations were performed using Statistical Analysis System® software, version 9.4, procedure power (SAS Institute Inc, Cary NC). Standardized effect sizes were estimated by dividing the expected mean differences $\mu(P_{\Delta V})$ by observed standard deviations $\sigma(P_{\Delta V})$ of mean differences:

$$d = \mu(P_{\Delta V}) / \sigma(P_{\Delta V}). \quad (5)$$

Sample size calculations were based on a 2-sided significance level (α) of 5%, and a power ($1 - \beta$) of 80%. Assuming a normal distribution and equal variances in 2 equally sized groups (control and intervention groups in a given future trial), the minimum required sample sizes per group for an independent 2-sample *t*-test were obtained. Then, a good approximation to calculate the minimum sample size N_{group} from the effect size *d* for an expected treatment effect *t* is given as follows [28]:

$$N_{group} > 15.7 / (d^*t)^2 + 0.96. \quad (6)$$

The minimum sample size was calculated for 20%, 30%, and 50% expected therapy effect, respectively.

2.6. Association to other markers of disease progression

For each HD patient, the cytosine-adenine-guanine (CAG) age product (CAP) was calculated as previously described

$$CAP = (CAG - L) * age / K \quad (7)$$

where *L* and *K* are constants. *L* is an estimate of the lower limit of the CAG expansion at which phenotypic expression of the effects of mutant huntingtin could be observed, and *K* is a normalizing constant; when *L* = 30 and *K* = 6.27, CAP will be equal to 100 at the subject's expected age of onset of motor symptoms [23,24].

The combined unified Huntington's disease rating scale (cUHDRS) [25] was calculated by

$$cUHDRS = [(TFC - 10.4) / 1.9 - (TMS - 29.7) / 14.9 + (SDMT - 28.4) / 11.3 + (SWR - 66.1) / 20.1] + 10 \quad (8)$$

where TFC is the total functional capacity, TMS denotes the total motor score, SDMT denotes the symbol digit modality test, and SWR denotes the stroop word reading.

Association of the CAP-score and the cUHDRS to annual volume changes of brain structures (analyzed from MRI data) was investigated by Pearson correlation; additionally, annual volume changes were correlated with longitudinal changes of cUHDRS.

3. Results

3.1. Volume changes and corresponding longitudinal effect sizes

Volumetric results for HD patients and controls are listed in Table 2, together with corresponding standardized effect sizes and required samples sizes for longitudinal clinical trials with 20%, 30% and 50% expected therapy effect, respectively. Nuisance variables such as age, sex, and study site were not accounted for since the groups (Table 1) showed no significant differences concerning age, gender and center-wise ratio of patients to controls (*t*-test for age in HD vs controls was 0.14, *t*-test for gender in HD vs controls was 0.54). In detail, volume changes of the whole brain, the cerebellum and major compartments like

GM, WM and CSF resulted in absolute standardized effect sizes < 0.9. Amongst the cerebral lobes, only the parietal lobes achieved a higher value (i.e., -1.05). Amongst CSF compartments, the lateral ventricles reached the highest value (0.92). Highest standardized effect sizes were obtained for striatal substructures such as caudate (-1.35), putamen (-0.84), nucleus accumbens (-0.91), and the striatum itself (-1.15), defined as summed-up volumes of the aforementioned substructures. Additional brain structures also showed volume loss in HD patients, i.e., the cerebellum, hippocampus, amygdala, brainstem, midbrain, pons, medulla, insula, pallidum, and thalamus; however, the absolute standardized effect sizes were smaller than those of the striatal structures (Table 2). Corrected *p*-values showed significant differences of annualized volume changes between HD patients and controls for the striatal structures, the parietal lobes, and the lateral ventricles. In the corresponding sample size calculations, the lowest numbers were achieved by volumetric results for the caudate nucleus (i.e., 36 subjects per group, for detecting a therapy effect of 50%). Further results for therapy effects of 20% and 30% are provided in Table 2.

The combined score C achieved an effect size of -1.60 and reduced the required sample sizes for detecting a therapy effect of 50% to 26 subjects per group (HD patients and controls).

3.2. Association of imaging-based atrophy marker to CAP-score and cUHDRS

Association of annualized volume changes of single structures revealed no significant correlations to the CAP score and to the cUHDRS. Only the combined score correlated significantly with the CAP score, $r = -0.26$, $p = 0.04$ (CAP score was available only for HD patients, thus $N = 59$). The combined score also correlated significantly with longitudinal changes of cUHDRS (over the time period of the three visits, available for all subjects, thus $N = 99$): $r = -0.33$, $p < 0.01$.

4. Discussion

Effect sizes of neuroimaging-based parameters have already been reported for HD patients, i.e., caudate atrophy and ventricular expansion [4,6], and image-based models of brain volume biomarker changes in HD provide insights into HD progression [26]. The focus of the present study was the investigation of longitudinal HD-associated regional atrophy with an unbiased (fully automated and therefore rater-independent) volumetric technique in order to calculate standardized longitudinal effect sizes with the aim of providing sample sizes for future clinical trials. To this end, the short-term changes in brain structures were measured by ABV on prospectively acquired 3D MRI datasets of HD patients and controls recruited within the framework of the PADDINGTON multicenter study, i.e., a well-characterized cohort of participants.

The effect sizes of the ABV approach were relatively high requiring for the best performing structure (caudate) only 36 data sets per treatment group to detect a 50% treatment effect. When comparing to similar studies [6], the effect sizes were 0.46 (TL), 1.05 (PL), and 0.62 (OL) for lobar volumes compared to 0.25 (TL), 0.38 (PL), and 0.51 (OL) for cortical thinning. However, a direct comparison of the results, although originating from the same study sample, is not possible since first, the neuroimaging analysis techniques differed and second, the subject samples differed in the final analysis sample. It has to be held that the current study was not designed for a direct comparison of different MRI post-processing methods.

A combined score of three non-overlapping structures/compartments, i.e., striatum, parietal lobes, and lateral ventricles, even allowed to reduce the required sample size to $N = 26$. The compilation of this combined score was based on the results of this study (i.e., all selected parameters individually achieved relatively high standardized effect sizes) and on the fact that they represent non-overlapping structures/compartments which can be regarded as surrogate markers of HD

progression in different brain regions, i.e., the striatal structures as the mainly affected subcortical structures, the parietal lobes as a marker for cortical atrophy, and the lateral ventricles as a global atrophy marker of the cerebrum. Furthermore, the combined score correlated significantly with the CAP score and also with longitudinal changes of the cUHDRS. We suggest that this imaging parameter may be considered as a surrogate marker of disease progression. Therefore, we propose this combined score of volume changes as an outcome measure for clinical trials aimed at determining disease modification in HD.

This approach to combine the longitudinal volume changes in different brain structures and compartments has not been attempted in any prior study in HD to the best of our knowledge. The observation that neuroimaging-based outcome measures might help to reduce sample sizes for clinical trials has already been described for other neurodegenerative diseases such as Alzheimer's disease [27]. ABV has recently been used for this purpose in progressive supranuclear palsy (PSP) when it was applied to the longitudinal 3D MRI from 99 PSP patients assigned to placebo in two clinical trials, and the highest standardized effect sizes were observed for atrophy in the midbrain, the frontal lobes, and the third ventricle; the combined analysis of these three compartments allowed to reduce the required sample size to detect a 50% treatment effect to 65% fewer patients than for the PSP rating scale total score [28].

The advantage of imaging above clinical parameters, which has already been previously reported for HD [4], is probably due to the generally high variability of clinical ratings which reduces the effect sizes. Combining several measures to one score has been suggested to further increase effect sizes [29] and to reduce the required sample sizes for clinical trials. Such a combination score also turned out to be the optimal approach in our study. Although imaging-based outcome measures are not accepted (yet) as primary endpoints in phase III clinical trials aiming at demonstration of clinical efficacy, they might still be used as interim read-out in longer efficacy studies or as secondary outcome measures in efficacy studies, as previously discussed [6].

This study has also limitations: The cohort represents a defined stage of HD. Therefore, the macrostructural neuroimaging readout over 15 months may not be indicative of longer term functional or clinical improvement and may not be suitable for all types of intervention, i.e., its utility may be dependent on the mechanism-of-action of the therapy, together with the time required for it to mediate an effect [6]. Nevertheless, as a strength of this technical approach, this neuroimaging measure has already been validated in HD [11] and is able to track the progression of pathological atrophy over short time intervals [16]. It may thus provide a valuable biological marker in the assessment of disease-modifying compounds. However, future studies may be implemented that extend the application of this marker to further stages of HD. As a further limitation, pooling of the data from both hemispheres did not allow for the analysis of lateralization effects in regional volume reductions which have been described in HD [30]; however, for the aims of this study, anatomical structures had to be addressed irrespective of hemispherical differences. Data heterogeneity due to scanning at different sites might be considered a limitation, but, for harmonization purposes, the multicentric data were age- and gender-matched for HD patients and controls with equal numbers of HD patients and controls across the four sites. Since the HD and the control group showed no significant differences concerning age, gender, and centerwise ratio of patients to controls, we did not correct for nuisance variables such as age, sex, and study site. All subjects were investigated at the same field strength of 3.0 T with similar spatial resolution; the differences in echo time (TE) and repetition time (TR) of the scanning protocols resulted in equalized image contrasts.

In conclusion, we propose a combined score of volumetric changes (striatum, parietal lobes, and lateral ventricles) as imaging read-out for potential disease-modifying clinical trials in HD. A result of this study are realistic subject numbers as a basis for the conceptualization of

future studies in HD; a validation of the current results in terms of effect size and subject numbers could be a side-result of such a future study. The decision to define these three regions/compartments for the combined score was data-driven, but was on the other hand in full agreement with the pathoneuroanatomy of HD.

The usage of this score requires the least number of patients for detecting biological evidence for efficacy to slow down disease progression, i.e., group sizes of 26 HD patients and 26 controls are sufficient to detect a 50% medication effect on regional brain volume loss over 15 months.

Acknowledgements

This work was supported by the European Union - PADDINGTON project, contract no HEALTH-F2-2010-261358.

The authors are grateful to Volkmar Glauche, MSc, group leader at the Freiburg Brain Imaging Center, University Medical Center Freiburg, Germany, for his help with implementing the Longitudinal Registration Toolbox of SPM12 in the processing pipeline of ABV. Furthermore, MRI datasets obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>) and from the International Consortium for Brain Mapping (ICBM) database (<http://www.loni.usc.edu/ICBM>) have been used to improve and test the methods of volumetric MRI analysis applied in this study.

References

- [1] C.A. Ross, E.H. Aylward, E.J. Wild, D.R. Langbehn, J.D. Long, J.H. Warner, R.I. Scapellato, B.R. Leavitt, J.C. Stout, J.S. Paulsen, R. Reilmann, P.G. Unschuld, A. Wexler, R.L. Margolis, S.J. Tabrizi, Huntington disease: natural history, biomarkers and prospects for therapeutics, *Nat. Rev. Neurol.* 10 (2014) 204–216.
- [2] S. Klöppel, S.M. Henley, N.Z. Hobbs, R.C. Wolf, J. Kassubek, S.J. Tabrizi, R.S. Frackowiak, Magnetic resonance imaging of Huntington's disease: preparing for clinical trials, *Neuroscience* 164 (2009) 205–219.
- [3] J.S. Paulsen, J.D. Long, C.A. Ross, D.L. Harrington, C.J. Erwin, J.K. Williams, H.J. Westervelt, H.J. Johnson, E.H. Aylward, Y. Zhang, H.J. Bockholt, R.A. Barker, Prediction of manifest Huntington's disease with clinical and imaging measures: a prospective observational study, *Lancet Neurol.* 13 (2014) 1193–1201.
- [4] S.J. Tabrizi, R. Reilmann, R.A. Roos, A. Durr, B. Leavitt, G. Owen, R. Jones, H. Johnson, D. Craufurd, S.L. Hicks, C. Kennard, B. Landwehrmeyer, J.C. Stout, B. Borowsky, R.I. Scapellato, C. Frost, D.R. Langbehn, Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data, *Lancet Neurol.* 11 (2012) 42–53.
- [5] S.J. Tabrizi, R.I. Scapellato, G. Owen, A. Durr, B.R. Leavitt, R.A. Roos, B. Borowsky, B. Landwehrmeyer, C. Frost, H. Johnson, D. Craufurd, R. Reilmann, J.C. Stout, D.R. Langbehn, Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data, *Lancet Neurol.* 12 (2013) 637–649.
- [6] N.Z. Hobbs, R.E. Farmer, E.M. Rees, J.H. Cole, S. Haider, I.B. Malone, R. Sprengelmeyer, H. Johnson, H.-P. Müller, S.D. Süsmuth, R.A. Roos, A. Durr, C. Frost, R.I. Scapellato, G.B. Landwehrmeyer, S.J. Tabrizi, Short-interval observational data to inform clinical trial design in Huntington's disease, *J. Neurol. Neurosurg. Psychiatry* 86 (2015) 1291–1298.
- [7] D.S. Majid, A.R. Aron, W. Thompson, S. Sheldon, S. Hamza, D. Stoffers, D. Holland, N.Z. Hobbs, J. Corey-Bloom, A.M. Dale, Basal ganglia atrophy in prodromal Huntington's disease is detectable over one year using automated segmentation, *Mov. Disord.* 26 (2011) 2544–2551.
- [8] J. Kassubek, F.D. Juengling, T. Kioschies, K. Henkel, J. Karitzky, B. Kramer, D. Ecker, J. Andrich, C. Saft, P. Kraus, A.J. Aschoff, A.C. Ludolph, G.B. Landwehrmeyer, Topography of cerebral atrophy in early Huntington's disease: a voxel based morphometric MRI study, *J. Neurol. Neurosurg. Psychiatry* 75 (2004) 213–220.
- [9] L.M. Ahveninen, J.C. Stout, N. Georgiou-Karistianis, V. Lorenzetti, Y. Glikmann-Johnston, Reduced amygdala volumes are related to motor and cognitive signs in Huntington's disease: the IMAGE-HD study, *Neuroimaging. Clin.* 18 (2018) 881–887.
- [10] H.J. Huppertz, J. Kröll-Seger, S. Klöppel, R.E. Ganz, J. Kassubek, Intra- and inter-scanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures, *Neuroimage* 49 (2010) 2216–2224.
- [11] J. Kassubek, E.H. Pinkhardt, A. Dietmaier, A.C. Ludolph, G.B. Landwehrmeyer, H.J. Huppertz, Fully automated atlas-based MR imaging volumetry in Huntington disease, compared with manual volumetry, *AJNR Am. J. Neuroradiol.* 32 (2011) 1328–1332.
- [12] H.-J. Huppertz, L. Möller, M. Südmeyer, R. Hilker, E. Hattingen, K. Egger, F. Amtage, G. Respondek, M. Stamelou, A. Schnitzler, E.H. Pinkhardt, W.H. Oertel, S. Knake, J. Kassubek, G.U. Höglinger, Differentiation of neurodegenerative Parkinsonian syndromes by volumetric MRI analysis and support vector machine classification, *Mov. Disord.* 31 (2016) 1506–1517.

- [13] R. Opfer, P. Suppa, T. Kepp, L. Spies, S. Schippling, H.J. Huppertz, Alzheimer's Disease Neuroimaging Initiative, Atlas based brain volumetry: how to distinguish regional volume changes due to biological or physiological effects from inherent noise of the methodology, *Magn. Reson. Imaging* 34 (2016) 455–461.
- [14] N.Z. Hobbs, J.H. Cole, R.E. Farmer, E.M. Rees, H.E. Crawford, I.B. Malone, R.A. Roos, R. Sprengelmeyer, A. Dürr, B. Landwehrmeyer, R.I. Scahill, S.J. Tabrizi, C. Frost, Evaluation of multi-modal, multi-site neuroimaging measures in Huntington's disease: baseline results from the PADDINGTON study, *Neuroimag. Clin.* 2 (2012) 204–211.
- [15] Huntington's Disease Study Group, Unified Huntington's disease rating scale: reliability and consistency, *Mov. Disord.* 11 (1996) 136–142.
- [16] L. Frings, I. Mader, B.G. Landwehrmeyer, C. Weiller, M. Hüll, H.J. Huppertz, Quantifying change in individual subjects affected by frontotemporal lobar degeneration using automated longitudinal MRI volumetry, *Hum. Brain Mapp.* 33 (2012) 1526–1535.
- [17] G.U. Höglinger, H.J. Huppertz, S. Wagenpfeil, M.V. Andrés, V. Belloch, T. León, T. Del Ser, TAUROS MRI Investigators, Tideglusib reduces progression of brain atrophy in progressive supranuclear palsy in a randomized trial, *Mov. Disord.* 29 (2014) 479–487.
- [18] J. Ashburner, G.R. Ridgway, Symmetric diffeomorphic modeling of longitudinal structural MRI, *Front. Neurosci.* 6 (2013) 197.
- [19] J.M. Goldstein, L.J. Seidman, N. Makris, T. Ahern, L.M. O'Brien, V.S. Caviness Jr., D.N. Kennedy, S.V. Faraone, M.T. Tsuang, Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability, *Biol. Psychiatry* 61 (2007) 935–945.
- [20] A. Hammers, R. Allom, M.J. Koeppe, S.L. Free, R. Myers, L. Lemieux, T.N. Mitchell, D.J. Brooks, J.S. Duncan, Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe, *Hum. Brain Mapp.* 19 (2003) 224–247.
- [21] D.W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K.L. Narr, R.A. Poldrack, R.M. Bilder, A.W. Toga, Construction of a 3D probabilistic atlas of human cortical structures, *Neuroimage* 39 (2008) 1064–1080.
- [22] I.B. Malone, K.K. Leung, S. Clegg, J. Barnes, J.L. Whitwell, J. Ashburner, N.C. Fox, G.R. Ridgway, Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance, *Neuroimage* 104 (2015) 366–372.
- [23] C.A. Ross, E.H. Aylward, E.J. Wild, D.R. Langbehn, J.D. Long, J.H. Warner, R.I. Scahill, B.R. Leavitt, J.C. Stout, J.S. Paulsen, R. Reilmann, P.G. Unschuld, A. Wexler, R.L. Margolis, S.J. Tabrizi, Huntington disease: natural history, biomarkers and prospects for therapeutics, *Nat. Rev. Neurol.* 10 (2014) 204–216.
- [24] J.H. Warner, C. Sampaio, Modeling variability in the progression of Huntington's disease a novel modeling approach applied to structural imaging markers from TRACK-HD, *CPT Pharmacometrics Syst. Pharmacol.* 5 (2016) 437–445.
- [25] S.A. Schobel, G. Palermo, P. Auinger, J.D. Long, S. Ma, O.S. Khwaja, D. Trundell, M. Cudkowicz, S. Hersch, C. Sampaio, E.R. Dorsey, B.R. Leavitt, K.D. Kieburtz, J.J. Sevigny, D.R. Langbehn, S.J. Tabrizi, TRACK-HD, COHORT, CARE-HD, 2CARE Huntington Study Group Investigators, Motor, cognitive, and functional declines contribute to a single progressive factor in early HD, *Neurology* 89 (2017) 2495–2502.
- [26] P.A. Wijeratne, A.L. Young, N.P. Oxtoby, R.V. Marinescu, N.C. Firth, E.B. Johnson, A. Mohan, C. Sampaio, R.I. Scahill, S.J. Tabrizi, D.C. Alexander, An image-based model of brain volume biomarker changes in Huntington's disease, *Ann. Clin. Transl. Neurol.* 5 (2018) 570–582.
- [27] D.M. Cash, C. Frost, L.O. Iheme, D. Ünay, M. Kandemir, J. Fripp, O. Salvado, P. Bourgeat, M. Reuter, B. Fischl, M. Lorenzi, G.B. Frisoni, X. Pennec, R.K. Pierson, J.L. Gunter, M.L. Senjem, C.R. Jack Jr., N. Guizard, V.S. Fonov, D.L. Collins, M. Modat, M.J. Cardoso, K.K. Leung, H. Wang, S.R. Das, P.A. Yushkevich, I.B. Malone, N.C. Fox, J.M. Schott, S. Ourselin, Assessing atrophy measurement techniques in dementia: results from the MIRIAD atrophy challenge, *Neuroimage* 123 (2015) 149–164.
- [28] G.U. Höglinger, J. Schöpe, M. Stamelou, J. Kassubek, T. Del Ser, A.L. Boxer, S. Wagenpfeil, H.J. Huppertz, AL-108-231 Investigators, Tauros MRI Investigators, Movement Disorder Society-Endorsed PSP Study Group, Longitudinal magnetic resonance imaging in progressive supranuclear palsy: a new combined score for clinical trials, *Mov. Disord.* 32 (2017) 842–852.
- [29] R. Jones, J.C. Stout, I. Labuschagne, M. Say, D. Justo, A. Coleman, E.M. Dumas, E. Hart, G. Owen, A. Durr, B.R. Leavitt, R. Roos, A. O'Regan, D. Langbehn, S.J. Tabrizi, C. Frost, The potential of composite cognitive scores for tracking progression in Huntington's disease, *J. Huntingtons Dis.* 3 (2014) 197–207.
- [30] L. Minkova, S. Gregory, R.I. Scahill, A. Abdulkadir, C.P. Kaller, J. Peter, J.D. Long, J.C. Stout, R. Reilmann, R.A. Roos, A. Durr, B.R. Leavitt, S.J. Tabrizi, S. Klöppel, TRACK-HD Investigators, Cross-sectional and longitudinal voxel-based grey matter asymmetries in Huntington's disease, *Neuroimag. Clin.* 17 (2017) 312–324.