Normative brain volume reports may improve differential diagnosis of dementing neurodegenerative diseases in clinical practice

Dennis M. Hedderich, Michael Dieckmeyer, Tiberiu Andrisan, Marion Ortner, Lioba Grundl, Simon Schön, Per Suppa, Tom Finck, et al.

European Radiology

ISSN 0938-7994

Eur Radiol
DOI 10.1007/s00330-019-06602-0

Springer

European Radiology
https://doi.org/10.1007/s00330-019-06602-0

NEURO

Normative brain volume reports may improve differential diagnosis of dementing neurodegenerative diseases in clinical practice

Dennis M. Hedderich[1] · Michael Dieckmeyer[1] · Tiberiu Andrisan[1] · Marion Ortner[2] · Lioba Grundl[1] · Simon Schön[1] · Per Suppa[3] · Tom Finck[1] · Kornelia Kreiser[1] · Claus Zimmer[1] · Igor Yakushev[4] · Timo Grimmer[2]

Received: 16 September 2019 / Revised: 23 October 2019 / Accepted: 27 November 2019

Abstract

Objectives Normative brain volume reports (NBVRs) are becoming more and more available for the workup of dementia patients in clinical routine. However, it is yet unknown how this information can be used in the radiological decision-making process. The present study investigates the diagnostic value of NBVRs for detection and differential diagnosis of distinct regional brain atrophy in several dementing neurodegenerative disorders.

Methods NBVRs were obtained for 81 consecutive patients with distinct dementing neurodegenerative diseases and 13 healthy controls (HC). Forty Alzheimer's disease (AD; 18 with dementia, 22 with mild cognitive impairment (MCI), 11 posterior cortical atrophy (PCA)), 20 frontotemporal dementia (FTD), and ten semantic dementia (SD) cases were analyzed, and reports were tested qualitatively for the representation of atrophy patterns. Gold standard diagnoses were based on the patients' clinical course, FDG-PET imaging, and/or cerebrospinal fluid (CSF) biomarkers following established diagnostic criteria. Diagnostic accuracy of pattern representations was calculated.

Results NBVRs improved the correct identification of patients vs. healthy controls based on structural MRI for rater 1 ($p < 0.001$) whereas the amount of correct classifications was rather unchanged for rater 2. Correct differential diagnosis of dementing neurodegenerative disorders was significantly improved for both rater 1 ($p = 0.001$) and rater 2 ($p = 0.022$). Furthermore, interrater reliability was improved from moderate to excellent for both detection and differential diagnosis of neurodegenerative diseases ($\kappa = 0.556/0.894$ and $\kappa = 0.403/0.850$, respectively).

Conclusion NBVRs deliver valuable and observer-independent information, which can improve differential diagnosis of neurodegenerative diseases.

Key Points

- *Normative brain volume reports increase detection of neurodegenerative atrophy patterns compared to visual reading alone.*
- *Differential diagnosis of regionally distinct atrophy patterns is improved.*
- *Agreement between radiologists is significantly improved from moderate to excellent when using normative brain volume reports.*

Keywords Alzheimer disease · Frontotemporal lobar degeneration · Magnetic resonance imaging · Neurodegenerative disorder · Brain

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06602-0) contains supplementary material, which is available to authorized users.

✉ Dennis M. Hedderich
dennis.hedderich@tum.de

[1] Department of Neuroradiology, Klinikum rechts der Isar, Technical University of Munich, School of Medicine, Ismaninger Str. 22, 81675 Munich, Germany

[2] Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, Technical University of Munich, School of Medicine, Munich, Germany

[3] Department of Nuclear Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany

[4] Department of Nuclear Medicine, Klinikum rechts der Isar, Technical University of Munich, School of Medicine, Munich, Germany

Published online: 30 January 2020

Abbreviations and acronyms

Aβ42	Beta amyloid 1–42
AD	Alzheimer's disease
CDR	Clinical Dementia Rating scale
CDR-SOB	Clinical Dementia Rating scale, Sum Of Boxes
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
CSF	Cerebrospinal fluid
DLB	Dementia with Lewy bodies
EADC	European Alzheimer's Disease Consortium
FDG-PET	Fluorodeoxyglucose positron emission tomography
FTD	Frontotemporal dementia
GCA	Global cortical atrophy
GM	Gray matter
HC	Healthy controls
LP	Lumbar puncture
MCI	Mild cognitive impairment
MMSE	Mini-Mental State Examination
MPRAGE	Magnetization prepared rapid gradient echo
MTA	Mesial temporal atrophy
NBVR	Normative brain volume report
NPV	Negative predictive value
PCA	Posterior cortical atrophy
PCC	Posterior cingulate cortex
PiB	Pittsburgh compound B
PPV	Positive predictive value
pTau	Phosphorylated Tau 181
SD	Semantic dementia
TE	Time to echo
TI	Time to inversion
TIV	Total intracranial volume
TPJ	Temporoparietal junction
TR	Time to repetition
tTau	Total Tau
WM	White matter

Introduction

Magnetic resonance imaging (MRI) plays a key role in the workup and differential diagnosis of neurodegenerative diseases [1, 2]. Since structural brain imaging is frequently performed early on in the diagnostic process, MRI has the potential to significantly guide or shorten a patient's diagnostic workup if a characteristic imaging pattern is observed or if pathological brain atrophy can be ruled out [1, 3]. To date, the radiologist in clinical routine mostly visually evaluates global and regional atrophy. This is notoriously difficult, dependent on the radiologist's level of expertise, and yields high intra- and interrater variability. Although visual rating scales have improved these issues and increased diagnostic accuracy, MRI-based image interpretation has probably not yet reached its full potential for evaluating regional cerebral atrophy [4, 5].

Technical advances have made it possible to integrate whole-brain volumetry into the clinical workflow and to obtain normative brain volume reports (NBVRs), which compare measured volumes of different brain structures with a healthy cohort after adjusting for sex and age [6, 7]. These NBVRs can present deviations from normal tissue volumes either as points plotted against a normal distribution and standard deviations or by color-coded, whole-brain statistical parametric maps (SPM). A similar approach has been introduced to fluorodeoxyglucose positron emission tomography (FDG-PET) imaging of the brain more than two decades ago [8]. Stereotactic surface projections have been shown to increase diagnostic accuracy in the workup of brain glucose metabolism [9].

NBVRs provide not only absolute volumes but also reference volumes of measured brain structures of a normative cohort of healthy individuals. While absolute volumetric measurements would be very hard to interpret, in our opinion, the addition of reference values and the visual representation of deviations from normal brain volumes as SPMs have great potential to facilitate the detection of deviations from normal aging and to make the identification of pathognomonic atrophy patterns easier. The purpose of this study is to evaluate the potential value of NBVRs in clinical practice with regard to identification and differential diagnosis of dementing neurodegenerative diseases with different regional atrophy patterns compared to visual inspection of atrophy patterns by neuroradiologists alone in a cohort of 94 consecutive patients with different neurodegenerative diseases and mixed with age-matched healthy controls.

Materials and methods

Study cohort and study design

Retrospective analysis of consecutive patients with distinct neurodegenerative diseases who were referred from the Centre for Cognitive Disorders of our hospital for imaging at our hybrid PET-MRI scanner (Siemens Biograph mMR PET-MRI, Siemens Healthineers) who fulfilled all inclusion criteria between 1 Jan. 2011 and 30 Sept. 2018. Retrospective inclusion of patients with Alzheimer's disease (AD) was stopped at 40 (22 at mild cognitive impairment (MCI) stage, 18 at dementia stage) in order to avoid largely inhomogeneous group sizes. Healthy controls (HC) served as control participants in former AD studies and underwent imaging at the same scanner. Besides available PET-MRI examination, the inclusion criteria for patients were sufficient quality of structural brain MRI including a 3D-T1 gradient echo sequence with a resolution of $1 \times 1 \times 1$ mm^3 (e.g., Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE)), an established diagnosis of sporadic AD (MCI) [10] or dementia stage [11]),

posterior cortical atrophy (PCA) [12], frontotemporal dementia (FTD), or semantic dementia (SD) [13] and absence of any concomitant neurological or psychiatric disorder. In this retrospective study, all patients were diagnosed according to then recent guidelines. Complying with current nomenclature, "FTD" corresponds to behavioral variant frontotemporal dementia (bvFTD) and SD corresponds to semantic variant primary progressive aphasia (svPPA). The reference standard diagnosis was based on expert diagnosis as a result of biomarker information (FDG-PET, amyloid PET, and/or cerebrospinal fluid (CSF) amyloid-β42, phosphorylated Tau (pTau) and total Tau (tTau)), clinical examination, cognitive testing, and clinical disease course. The diagnostic workup comprised thorough clinical and neuropsychological testing including the Mini-Mental State Examination (MMSE) and the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological assessment battery. Severity of dementia was rated using the global score of the Clinical Dementia Rating scale (CDR global).

Detailed information about patient characteristics and their clinical workup is given in Table 1.

Image acquisition and analysis

Three-dimensional, T1-weighted MRI scans of patients and healthy individuals of a normative database from our hybrid PET-MRI platform were exported from the local Picture Archiving and Communication System (PACS). The normative database comprised 26 healthy subjects with a mean age of 57 years (standard deviation of 11 years) ranging from 41 to 81 years. These healthy individuals served as normal controls in previous prospective trials and were scanned using the same scanner and sequence specifications as the cohort evaluated in this study [14]. According to previous studies, statistical power for asymmetric two-sample *t* tests can be considered sufficient if the normative database comprises 20–30 subjects [15, 16]. Scan parameters were as follows: time to repetition (TR) = 2300 ms, time to echo (TE) = 2.98 ms, time to inversion (TI) = 900 ms, flip angle = 9°, acquisition matrix = 256 × 240 mm^2, and voxel size = 1 × 1 × 1 mm^3. MRI scans were processed using the Biometrica analysis platform (jung diagnostics GmbH), which incorporates all the subsequent postprocessing steps. The T1 MRI images were segmented using a previously described and validated atlas-based volumetry approach implemented in SPM12 [17, 18]. In brief, MRI brain scans were segmented into tissue class component images representing either gray matter (GM), white matter (WM), or CSF. The total intracranial volume (TIV) was estimated using a method which was recently introduced and validated by Malone et al [19]. Results of the tissue segmentation were visually checked for segmentation errors. All tissue segmentations passed quality control. Hereafter, standard voxel-based morphometry (VBM) [20] as provided by the SPM12 software package was applied to the individual GM tissue class component image of a patient using a modification of Mühlau et al [21] for asymmetric statistical designs. Spatial correspondence between the individual GM tissue class component image of the patient and the GM tissue class component images of the normative database was established via a high-dimensional nonlinear image registration technique (DARTEL) [22]. GM volumes on voxel level were adjusted for TIV and age to minimize the impact of these confounding variables on statistical analysis. The adjustment was performed by computing the residuals from a bilinear regression function. Voxel-wise *t* tests of the age and TIV adjusted GM volumes between patients and healthy individuals were performed. An extend threshold of 125 voxels corresponding to a cluster volume of 1 ml was set to partially correct for multiple comparisons [23]. The resulting *p* values were presented as color-coded overlay on axial slides and surface projections. For illustration of image postprocessing and NBVR generation, please see Fig. 1.

Brain MRI and NBVR evaluation

All brain MRI images and NBVRs were evaluated by two neuroradiologists with 2 years and 7 years of experience, respectively. Additional information available to the reviewers were sex and age; ratings were performed blinded

Table 1 Study participant characteristics

	n	Sex (% male)	Age (mean ± StD)	MMSE (mean ± StD)	CDR global (median (min-max))
HC	13	46.2%	66.2 ± 9.2	28.2 ± 1.4	0.0 (0.0–0.0)
AD	18	44.4%	66.2 ± 8.2	16.9 ± 6.5	2.0 (1.0–3.0)
MCI AD	22	63.6%	68.6 ± 6.5	24.0 ± 4.2	0.5 (0.5–1.0)
PCA	11	27.3%	63.3 ± 7.6	21.7 ± 4.4	1.0 (1.0–2.0)
FTD	20	25.0%	62.3 ± 8.7	21.9 ± 6.7	1.0 (1.0–2.0)
SD	10	50.0%	66.4 ± 7.3	23.3 ± 3.0	1.0 (1.0–1.0)

AD Alzheimer's disease; *MCI AD* mild cognitive impairment due to Alzheimer's disease; *CDR global* clinical dementia rating scale, global score; *FTD* frontotemporal dementia; *HC* healthy control; *min* minimum; max maximum; *MMSE* Mini-Mental State examination; *PCA* posterior cortical atrophy; *SD* semantic dementia; *StD* standard deviation

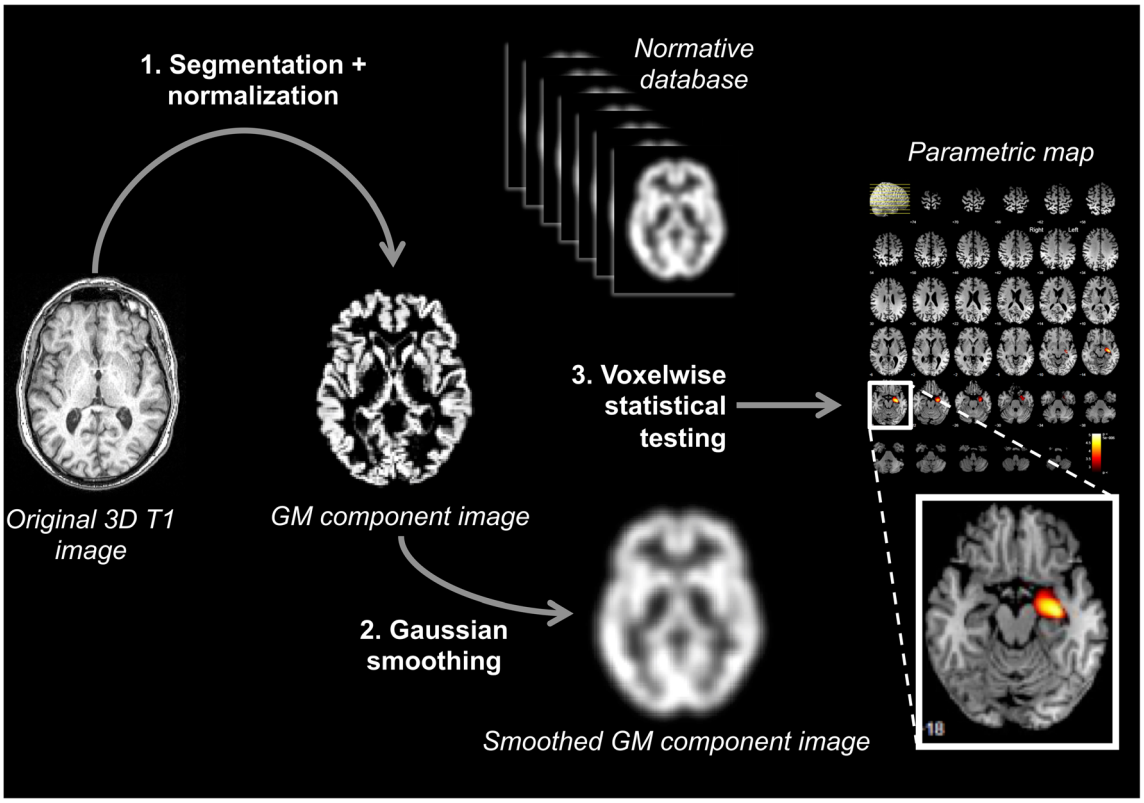

Fig. 1 Image processing pipeline and normative brain volume report generation. The processing pipeline incorporates classical voxel-based morphometry: the original 3D T1-image (MPRAGE) is segmented and normalized into a gray matter (GM) component image (first processing step); hereafter, the normalized GM tissue class component image is smoothed using a Gaussian filter (second step). After smoothing, a voxel-based statistical test of the individual smoothed GM component image against a normative database is carried out and produces a parametric map. The parametric map indicates statistically significant reductions of gray matter on voxel level (in color). Abbreviations: GM, gray matter; MPRAGE, Magnetization Prepared Rapid Gradient Echo

for all other clinical or biomarker information. Visual assessment of regional brain atrophy was based on axial, coronal, and sagittal reconstructions of the 3D T1-weighted MRI sequence at 1 mm^3 isotropic resolution. The raters, if needed, could adapt image reconstructions. Raters were not aware of the distribution of diagnoses within the study cohort. The endpoint was to identify or exclude a neurodegenerative pattern of brain atrophy. Evaluation took place in two reading sessions. All brain MRI scans were evaluated both with and without an NBVR in two reading sessions by both raters separately. The order of the two types of evaluation was assigned randomly to exclude training effects. The two reading sessions were scheduled 4 weeks apart, in order to exclude a memory bias. Raters did not receive a study-specific training to assess brain regional atrophy patterns due to their strong clinical background in neuroradiology and to resemble clinical routine. Both raters had to state (I) whether there is abnormal brain volume loss present, suggestive of any neurodegenerative disease, and (II) whether the atrophy pattern allows making a differential diagnosis with respect to the exact disorder (AD, PCA, SD, FTD). For detection and differential diagnosis of brain atrophy patterns, the readers interpreted the SPMs of GM volume deviations from the normal control cohort, presented as both axial slices and 3D renderings at $p < 0.005$, uncorrected (see Fig. 2 for exemplary 3D renderings).

In order to simulate the decision-making process in clinical routine, the expected pathognomonic atrophy patterns were derived from the literature and defined as follows [24]:

AD type—symmetric or asymmetric atrophy of the medial temporal lobe, temporoparietal junction (TPJ), and posterior cingulate cortex (PCC). Frontal atrophy was facultative [25].

PCA type—symmetric or asymmetric atrophy in the posterior temporal lobe and the parietal lobe with prominent involvement of the occipital lobe [26].

FTD type—symmetric or asymmetric atrophy in the frontal and anterior temporal lobes, with predominance of the frontal lobes or similar atrophy of frontal and anterior temporal lobes [27].

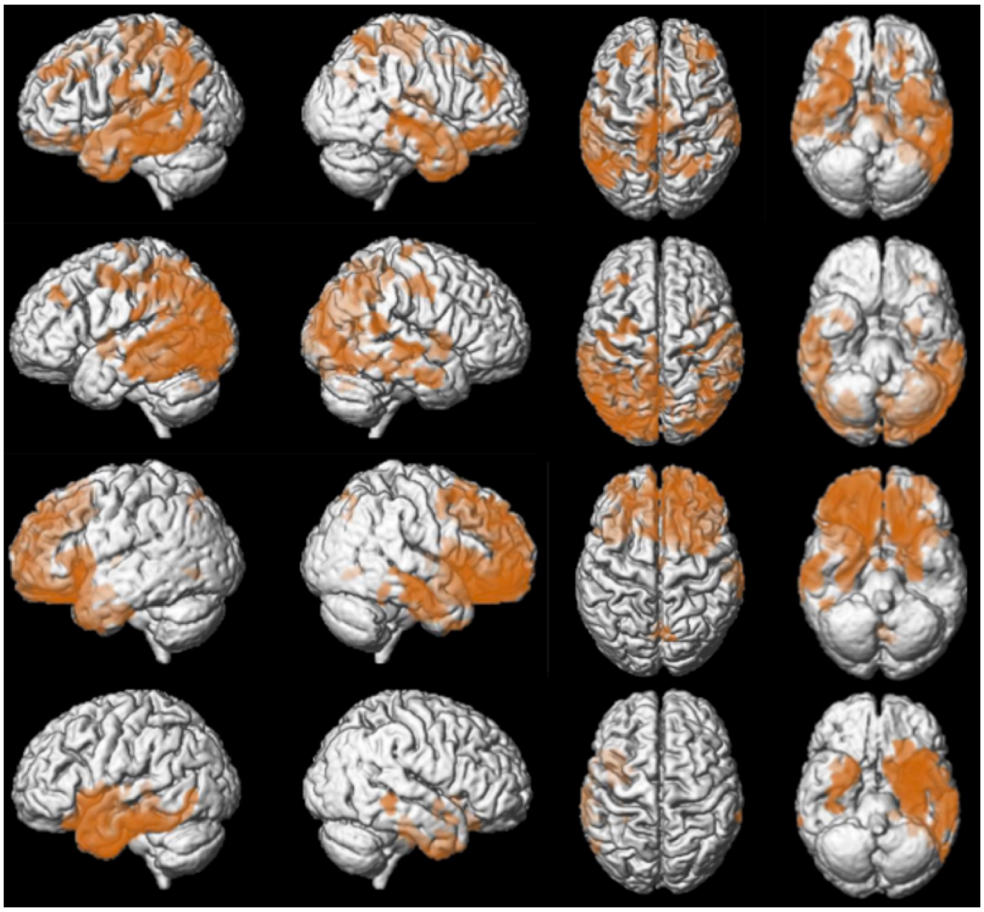

Fig. 2 Exemplary patterns of neurodegeneration for AD (top row), PCA (2nd row), FTD (3rd row), and SD (bottom row). Examples are shown as 3D renderings of age-corrected GMV reductions (orange) at $p < 0.005$ uncorrected. Abbreviations: AD, Alzheimer's disease; FTD, frontotemporal dementia; GMV, gray matter volume; PCA, posterior cortical atrophy; SD, semantic dementia

SD type—symmetric or asymmetric atrophy in the frontal and anterior temporal lobes, with predominance of the temporal lobes [28].

Exemplary AD-type, PCA-type, FTD-type, and SD-type patterns of neurodegeneration are depicted in Fig. 2.

Statistical analysis

Individual rating results were analyzed using crosstables. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic accuracy were calculated. Differences of intra-individual correct classifications between visual inspection alone and presence of a NBVR were calculated using McNemar's test. Differences of patient characteristics were calculated using either two-sided *t* test (age), Mann-Whitney *U* tests (MMSE), or chi-square test (sex). Differences were considered statistically significant for $p < 0.05$. To assess interrater reliability, Cohen's kappa was used. All statistical tests were performed using SPSS version 25.0 (SPSS, IBM Corp. 2017).

Ethics statement

Study procedures, namely the anonymous analysis of the MRI data, were approved by the ethics committee of the Technical University of Munich (Reference # 176/18s). Written informed consent was waived due to the retrospective nature of this analysis.

Results

Study cohort

The study cohort consisted of 13 HCs and 81 consecutive patients with distinct dementing neurodegenerative diseases: 40 AD; 18 at dementia stage, 22 at MCI stage), 11 PCA, 20 FTD, and ten SD cases. There were no significant differences between all patients and HCs regarding age ($p = 0.515$) and sex ($p = 0.843$). No unrelated gross pathology was detected on brain MRIs. Patients suffering from any neurodegenerative disease exhibited significantly lower MMSE scores than HCs (21.5 ± 5.8 vs 28.2 ± 1.4; $p < 0.001$). All individuals in

the HC group were rated as 0 on the CDR global scale. The median score on the Fazekas scale was not statistically different ($p = 0.067$) between all patients (median 1, range 0–3) and HCs (median 1, range 0–1). Detailed demographic characteristics are given in Table 1.

Diagnostic accuracy for detection of neurodegenerative diseases

For detection of any neurodegenerative disease, conventional visual inspection of brain MRI yielded sensitivities and specificities of 0.67 [95% CI 0.55–0.77]/0.85 [0.54–0.97] and 0.84 [0.74–0.91]/0.92 [0.62–1.00] for raters 1 and 2, respectively. Detection of any neurodegenerative pattern based on visual evaluation of NBVRs yielded sensitivities and specificities of 0.86 [0.77–0.93]/1.00 [0.72–1.00] and 0.81 [0.71–0.89]/1.00 [0.72–1.00], respectively. When comparing correctly classified individuals (neurodegenerative disease vs. healthy controls), the assessment supported by NBVRs led to a significantly increased amount of correct diagnoses in rater 1 (visual inspection alone, 65 of 94 correct classifications vs. NBVR, 83 of 94 correct classifications; $p < 0.001$) whereas the correct classifications by rater 2 remained largely unchanged (visual inspection alone, 80 of 94 correct classifications vs. NBVR, 79 of 94 correct classifications; $p = 1.00$). Crosstable analyses can be found in Table 2; additional metrics of diagnostic accuracy can be found in Table 3. In order to illustrate our results, two examples of patients with AD who were initially judged false negative as normal aging by visual inspection alone and classified correctly as AD with NBVRs.

Diagnostic accuracy for differential diagnosis of neurodegenerative diseases

Diagnostic accuracy for the differential diagnosis of neurodegenerative diseases was calculated separately for the four included entities AD, PCA, FTD, and SD. Most importantly, sensitivity and specificity improved for AD for rater 1 (visual inspection alone, 0.66 [95% CI, 0.43–0.85]/0.57 [0.40–0.72]; NBVR, 0.93 [0.66–1.00]/0.91 [0.74–0.98]). For rater 2, specificity increased for detection of an AD-type pattern using NBVRs (visual inspection alone, 0.86 [0.66–0.95]/0.55 [0.39–0.70]; NBVR, 0.83 [0.62–0.95]/0.91 [0.75–0.98]). Using NBVRs substantially increased the sensitivity to detect a pattern of atrophy for FTD and SD in both raters: FTD-type pattern (rater 1: visual inspection alone, 0.53 [0.29–0.76]/1.00 [0.89–1.00]; NBVR, 0.94 [0.68–1.00]/0.97 [0.81–1.00]; rater 2: visual inspection alone, 0.37 [0.17–0.61]/0.98 [0.88–1.00]; NBVR, 0.82 [0.56–0.95]/1.00 [0.89–1.00]), SD-type pattern (rater 1: visual inspection alone, 0.40 [0.14–0.73]/1.00 [0.91–1.00]; NBVR, 0.90 [0.54–1.00]/0.97 [0.84–1.00]; rater 2: visual inspection alone, 0.30 [0.08–0.65]/0.98 [0.90–1.00]; NBVR, 0.90 [0.54–0.99]/0.92 [0.79–0.97]).

There was no atrophy pattern discernible at all on NBVRs for rater 1 in 11 patients (9 AD, 1 PCA, 1 FTD) and for rater 2 in 15 patients (12 AD, 2 PCA, 1 FTD).

When comparing correctly classified individuals regarding their differential diagnoses of neurodegenerative disease, the amount of correct diagnoses increased significantly for both rater 1 (visual inspection alone, 33 of 94 correct classifications vs. NBVR, 55 of 94 correct classifications; $p = 0.001$) and rater 2 (visual inspection alone, 44 of 94 correct classifications vs. NBVR, 61 of 94 correct classifications; $p = 0.022$). Crosstable analyses can be found in supplemental Table S1. For additional metrics of differential diagnostic accuracy, please see Table 3.

Interrater reliability

Cohen's κ was calculated for interrater agreement with respect to (1) detection of any neurodegenerative disease pattern and (2) differential diagnosis of included neurodegenerative disease entities AD, PCA, FTD, and SD. Interrater agreement increased substantially from moderate to excellent if an NBVR was present for both detection of any neurodegenerative disease

Table 2 Crosstable analysis for differentiation of healthy controls (HC) from patients with any neurodegenerative disease (ND). Numbers are given for raters 1 and 2 for visual inspection only and NBVR evaluation

		Rater 1				Rater 2			
		Visual inspection		NBVR evaluation		Visual inspection		NBVR evaluation	
		HC	ND	HC	ND	HC	ND	HC	ND
Reference standard	HC	11	2	13	0	12	1	13	0
	ND	27	54	11	70	13	68	15	66

HC healthy control, *NBVR* normative brain volume report, *ND* neurodegenerative disease

Table 3 Indicators of diagnostic accuracy of classification of neurodegenerative patterns for visual inspection only/NBVR evaluation

	Rater 1				Rater 2			
	Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV
HC vs. any neurodegenerative disorder	0.66/0.86	0.85/1.0	0.96/1.00	0.29/0.54	0.84/0.81	0.92/1.00	0.99/1.00	0.48/0.46
AD	0.66/0.93	0.57/0.91	0.47/0.82	0.75/0.97	0.86/0.83	0.55/0.91	0.56/0.87	0.85/0.89
PCA	0.75/0.67	0.82/1.00	0.4/1.00	0.95/0.95	0.91/0.71	0.93/0.94	0.71/0.63	0.98/0.96
FTD	0.53/0.94	1.00/0.97	1.00/0.94	0.84/0.97	0.37/0.82	0.98/1.00	0.88/1.00	0.81/0.93
SD	0.40/0.90	1.00/0.97	1.00/0.90	0.89/0.97	0.30/0.90	0.98/0.92	0.75/0.69	0.89/0.98

Indicators of diagnostic accuracy are given for visual inspection only/NBVR evaluation

AD Alzheimer's disease, *FTD* frontotemporal dementia, *HC* healthy control, *NPV* negative predictive value, *PCA* posterior cortical atrophy, *PPV* positive predictive value, *SD* semantic dementia, *Sens* sensitivity, *Spec* specificity

(visual inspection alone $\kappa = 0.556$ [95% CI, 0.387–0.725], NBVR evaluation $\kappa = 0.894$ [0.792–0.996]) and the differential diagnosis of neurodegenerative disease patterns (visual inspection alone $\kappa = 0.403$ [0.195–0.611], NBVR evaluation $\kappa = 0.850$ [0.727–0.973]).

Discussion

Investigating the diagnostic process of neurodegenerative disorders with structural MRI, we have shown a significantly improved detection of neurodegenerative atrophy patterns using NBVRs compared to visual evaluation only. This translated into significantly more correct differential diagnoses as well as substantially increased interrater agreement and thus underlines the potential of NBVRs to improve and standardize the workup of dementia patients with structural MRI.

Currently, further technological advances like automated or semi-automated diagnoses based on deep learning algorithms are being heavily discussed within the neuroradiological community [29, 30]. We have shown that good visual representation of regional relative brain volume deficits alone can improve diagnostic accuracy for neurodegenerative diseases, without diagnoses suggested by an algorithm. However, neuroradiologists have been hesitant to even adopt only more advanced image preprocessing and presentation techniques similar to stereotactic surface projections and NBVRs.

We have shown that identifying patterns of neurodegenerative disease based on NBVRs leads to more accurate diagnoses in patients with AD through increased sensitivity. This could be important in clinical practice in order to guide the diagnostic process. For example, if a pathognomonic pattern can be discerned by MRI-based brain volumetry, it could be reasonable to refer the patient directly to PET imaging for cerebral amyloid deposition and perform FDG-PET only in dubious cases.

Apart from superior overall differential diagnosis with NBVRs, we also observed a substantial increase in sensitivity for the detection of patients with FTD and SD. FTD and SD represent diseases of the frontotemporal lobar degeneration spectrum and are less prevalent than for example AD, which may add to the low sensitivity observed by visual evaluation alone [31]. NBRVs have the potential to visualize atrophy patterns in a very prominent way, thus attracting the radiologist's attention towards less frequent differential diagnoses such as FTD or SD. However, some reports from patients with neurodegenerative diseases did not show pathological atrophy patterns. This emphasizes that a negative NBVR does not exclude neurodegenerative disease and further workup may be necessary. This holds true at least for the commercially available product used in this publication, which is based on standard SPM algorithms for voxel-based morphometry (VBM) analyses. Future research will be needed to establish optimal preprocessing and thresholding parameters of structural MRI data to be used in clinical routine for detection and exclusion of neurodegenerative disorders. Also, additional indicators of neurodegeneration, e.g., reduced cortical thickness or reduced cortical complexity, might further increase the sensitivity of neurodegenerative disorder detection in clinical routine [32, 33]. However, no commercially available products adapted to clinical workflows allow for the analysis of these more sophisticated parameters of neurodegeneration.

One of the main advantages we have observed regarding the use of NBVRs is the superior interrater agreement. Interrater agreement is notoriously poor for mere visual evaluation of atrophy patterns [34]. In the past, semi-quantitative visual rating scales—most prominently the mesial temporal atrophy (MTA) score—have been shown to increase interrater reliability [4, 35, 36]. However, NBVRs can still improve the agreement among different radiologists with different levels of

experience and thus contribute to a more objective imaging report.

Several limitations have to be mentioned in this retrospective analysis of diagnostic accuracy. This is a first study from a single institution and a single MRI scanner and a rather small sample size. In our opinion, future studies are needed to test the generalizability of our results. For most neurodegenerative disorders, a definite diagnosis is made by the pathologist post mortem, which is usually lacking for patients in clinical routine and was also not available as reference standard in our study. We aimed at the highest possible level of confidence regarding the patients' diagnoses by integrating clinical evaluation based on established diagnostic criteria [10–13] including follow-up examinations, neuropsychological testing as well as PET and/or CSF biomarkers into the diagnosis. We retrospectively included a consecutive set of patients with four different neurodegenerative diseases with distinct atrophy patterns in order to test the ability of NBVRs to separate those regional differences. Since patients with dementia with Lewy bodies (DLB) usually do not show typical brain atrophy patterns, we did not include this entity in our cohort, which limits the generalizability of our results to a clinical setting. Also, raters might achieve higher diagnostic accuracies in study settings with preselected patient cohorts compared to the clinical setting. However, in the clinical setting, a patient's chief complaint might influence the neuroradiologist's decision towards a differential diagnosis, which was excluded here by blinding the raters for clinical information. Additionally, the inclusion criterion of a definite clinical diagnosis led to the exclusion of a substantial amount of patients, so that the prevalences of neurodegenerative diseases in our study cannot be considered representative of patients in a clinical setting and the high standard of diagnostic confidence might have introduced a selection bias. This unbalanced distribution of patients with neurodegenerative disease and normal controls might have an impact on ratings with regard to the identification of patients with any neurodegenerative disease due to a high pretest probability of study participants belonging to the patient group. However, raters were not aware of the distribution of diagnoses within the study cohort, and ratings aiming at the differentiation of regional atrophy patterns should not be affected as much due to the more balanced distribution of different entities within the patient group. When comparing MRI-based brain volumes of single subjects to a cohort of healthy controls, the size of the control cohort contributing to the normative database and between-scanner effects are crucial. In this work, we only included HCs scanned on the same scanner using the same specifications in the normative database. Consequently, the total number of subjects contributing to the normative database could be considered rather low. However, previous work has shown that the statistical power of an asymmetric two-sample t test starts to plateau after including 20 subjects into the normative database and very little improvement of statistical power is observed for more than 30 healthy controls [16].

In conclusion, we showed improved diagnostic accuracy regarding patterns of cerebral atrophy both for the presence of any neurodegenerative disorder in general and for the differential diagnoses for four entities with distinct regional atrophy patterns by use of NBVRs. Moreover, NBVRs significantly improved interrater agreement and thus could lead to a more precise and objective evaluation of patients with neurodegenerative diseases in clinical practice. Our results motivate further studies in larger cohorts and thorough investigation of the impact of NBVRs on workflow in radiological practices.

Acknowledgments We thank Dr. Lothar Spies and jung diagnostics GmbH for providing volumetric reports for the investigated MRI scans.

Funding information The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Timo Grimmer, MD.

Conflict of interest The authors of this manuscript declare relationships with the following companies: jung diagnostics GmbH, Hamburg, Germany. Jung diagnostics GmbH provided the volumetric reports for the current study. No further relationships exist between the authors and jung diagnostics and no conflict of interest is present. Per Suppa, MD was an employee of jung diagnostics until November 2018.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained (Reference # 176/18s).

Methodology
- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

References

1. Teipel S, Drzezga A, Grothe MJ et al (2015) Multimodal imaging in Alzheimer's disease: validity and usefulness for early detection. Lancet Neurol 14:1037–1053
2. Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM (2010) The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 6:67–77
3. Teipel S, Kilimann I, Thyrian JR, Klöppel S, Hoffmann W (2017) Potential role of neuroimaging markers for early diagnosis of dementia in primary care. Curr Alzheimer Res 15:18–27

4. Harper L, Fumagalli GG, Barkhof F et al (2016) MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. Brain 139:1211–1225
5. Wahlund LO, Westman E, van Westen D et al (2017) Imaging biomarkers of dementia: recommended visual rating scales with teaching cases. Insights Imaging 8:79–90
6. Potvin O, Dieumegarde L, Duchesne S; Alzheimer's Disease Neuroimaging Initiative (2017) Normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain. Neuroimage 156:315–339
7. Bruun M, Frederiksen KS, Rhodius-Meester HFM et al (2019) Impact of a clinical decision support tool on prediction of progression in early-stage dementia: a prospective validation study. Alzheimers Res Ther 16:91–101
8. Minoshima S, Frey KA, Koeppe RA, Foster NL, Kuhl DE (1995) A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET. J Nucl Med 36:1238–1248
9. Brown RKJ, Bohnen NI, Wong KK, Minoshima S, Frey KA (2014) Brain PET in suspected dementia: patterns of altered FDG metabolism. Radiographics 34:684–701
10. Albert MS, DeKosky ST, Dickson D et al (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7:270–279
11. McKhann GM, Knopman DS, Chertkow H et al (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7:263–269
12. Mendez MF, Ghajarania M, Perryman KM (2002) Posterior cortical atrophy: clinical characteristics and differences compared to Alzheimer's disease. Dement Geriatr Cogn Disord 14:33–40
13. Neary D, Snowden JS, Gustafson L et al (1998) Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 51:1546–1554
14. Savio A, Funger S, Tahmasian M et al (2017) Resting-state networks as simultaneously measured with functional MRI and PET. J Nucl Med 58:1314–1317
15. Chen W-P, Samuraki M, Yanase D et al (2008) Effect of sample size for normal database on diagnostic performance of brain FDG PET for the detection of Alzheimer's disease using automated image analysis. Nucl Med Commun 29:270–276
16. Buchert R (2008) On the effect of sample size of the normal database on statistical power of single subject analysis. Nucl Med Commun 29:837
17. Huppertz H-J, Kroll-Seger J, Kloppel S, Ganz RE, Kassubek J (2010) Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. Neuroimage 49:2216–2224
18. Opfer R, Suppa P, Kepp T, Spies L, Schippling S, Huppertz H-J (2016) 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 34:455–461
19. Malone IB, Leung KK, Clegg S et al (2015) Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance. Neuroimage 104:366–372
20. Ashburner J, Friston KJ (2000) Voxel-based morphometry - the methods. Neuroimage 11:805–821
21. Muhlau M, Wohlschlager AM, Gaser C et al (2009) Voxel-based morphometry in individual patients: a pilot study in early Huntington disease. AJNR Am J Neuroradiol 30:539–543
22. Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113
23. Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33:636–647
24. Risacher S, Saykin A (2013) Neuroimaging biomarkers of neurodegenerative diseases and dementia. Semin Neurol 33:386–416
25. Whitwell JL, Jack CR Jr, Przybelski SA et al (2011) Temporoparietal atrophy: a marker of AD pathology independent of clinical diagnosis. Neurobiol Aging 32:1531–1541
26. Lehmann M, Crutch SJ, Ridgway GR et al (2011) Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. Neurobiol Aging 32:1466–1476
27. Rohrer JD (2012) Structural brain imaging in frontotemporal dementia. Biochim Biophys Acta 1822:325–332
28. Rohrer JD, Warren JD, Modat M et al (2009) Patterns of cortical thinning in the language variants of frontotemporal lobar degeneration. Neurology 72:1562–1569
29. Lee G, Nho K, Kang B, Sohn KA, Kim D; for Alzheimer's Disease Neuroimaging Initiative (2019) Predicting Alzheimer's disease progression using multi-modal deep learning approach. Sci Rep 9:1952
30. Basaia S, Agosta F, Wagner L et al (2019) Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage Clin 21:101645
31. Diehl-Schmid J, Onur OA, Kuhn J, Gruppe T, Drzezga A (2014) Imaging frontotemporal lobar degeneration. Curr Neurol Neurosci Rep 14:1–11
32. King RD, Brown B, Hwang M, Jeon T, George AT (2010) Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease. Neuroimage 53:471–479
33. Cho Y, Seong J-K, Jeong Y, Shin SY (2012) Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage 59:2217–2230
34. Klöppel S, Yang S, Kellner E et al (2018) Voxel-wise deviations from healthy aging for the detection of region-specific atrophy. NeuroImage Clin 20:851–860
35. Scheltens P, Leys D, Barkhof F et al (1992) Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatry 55:967–972
36. Ferreira D, Cavallin L, Larsson EM et al (2015) Practical cut-offs for visual rating scales of medial temporal, frontal and posterior atrophy in Alzheimer's disease and mild cognitive impairment. J Intern Med 278:277–290