



# A Fully Automated Pipeline for Normative Atrophy in Patients with Neurodegenerative Disease

Christian Rummel<sup>1\*</sup>, Fabian Aschwanden<sup>1</sup>, Richard McKinley<sup>1</sup>, Franca Wagner<sup>1</sup>, Anke Salmen<sup>2</sup>, Andrew Chan<sup>2</sup> and Roland Wiest<sup>1</sup>

<sup>1</sup> Support Center for Advanced Neuroimaging (SCAN), University Institute for Diagnostic and Interventional Neuroradiology, Inselspital Bern, University of Bern, Bern, Switzerland, <sup>2</sup> Department of Neurology, Inselspital Bern, University of Bern, Bern, Switzerland

**Introduction:** Volumetric image analysis to detect progressive brain tissue loss in patients with multiple sclerosis (MS) has recently been suggested as a promising marker for “no evidence of disease activity.” Software packages for longitudinal whole-brain volume analysis in individual patients are already in clinical use; however, most of these methods have omitted region-based analysis. Here, we suggest a fully automatic analysis pipeline based on the free software packages FSL and FreeSurfer.

**Materials and methods:** Fifty-five T1-weighted magnetic resonance imaging (MRI) datasets of five patients with confirmed relapsing–remitting MS and mild to moderate disability were longitudinally analyzed compared to a morphometric reference database of 323 healthy controls (HCs). After lesion filling, the volumes of brain segmentations and morphometric parameters of cortical parcellations were automatically screened for global and regional abnormalities. Error margins and artifact probabilities of regional morphometric parameters were estimated. Linear models were fitted to the series of follow-up MRIs and checked for consistency with cross-sectional aging in HCs.

**Results:** As compared to leave-one-out cross-validation in a subset of the control dataset, anomaly detection rates were highly elevated in MRIs of two patients. We detected progressive volume changes that were stronger than expected compared to normal aging in 4/5 patients. In individual patients, we also identified stronger than expected regional decreases of subcortical gray matter, of cortical thickness, and areas of reducing gray–white contrast over time.

**Conclusion:** Statistical comparison with a large normative database may provide complementary and rater independent quantitative information about regional morphological changes related to disease progression or drug-related disease modification in individual patients. Regional volume loss may also be detected in clinically stable patients.

**Keywords:** structural magnetic resonance imaging, automated morphometry, multiple sclerosis, atrophy progression, individualized medicine

**Abbreviations:** CIS, clinically isolated syndrome; CSF, cerebrospinal fluid; EDSS, expanded disability status scale; eTIV, estimated total intracranial volume; FDR, false discovery rate; FLAIR, fluid-attenuated inversion recovery; GM, gray matter; HC, healthy control; LOOCV, leave-one-out cross-validation; MRI, magnetic resonance imaging; MS, multiple sclerosis; NEDA, no evidence of disease activity; RR-MS, relapsing–remitting multiple sclerosis; SBA, surface-based analysis; T1w, T1-weighted; VBM, voxel-based morphometry; WM, white matter.

## OPEN ACCESS

### Edited by:

Hans-Peter Müller,  
University of Ulm, Germany

### Reviewed by:

Jeroen J. G. Geurts,  
VU University Medical Center,  
Netherlands  
Patrizia Pantano,  
Sapienza Università di Roma, Italy

### \*Correspondence:

Christian Rummel  
crummel@web.de

### Specialty section:

This article was submitted to Applied  
Neuroimaging,  
a section of the journal  
Frontiers in Neurology

Received: 23 June 2017

Accepted: 18 December 2017

Published: 24 January 2018

### Citation:

Rummel C, Aschwanden F,  
McKinley R, Wagner F, Salmen A,  
Chan A and Wiest R (2018) A Fully  
Automated Pipeline for Normative  
Atrophy in Patients with  
Neurodegenerative Disease.  
Front. Neurol. 8:727.  
doi: 10.3389/fneur.2017.00727

## INTRODUCTION

Recently, the “no-evidence of disease activity” criteria (NEDA-3) have been extended (NEDA-4) in order to add cerebral atrophy as a potential surrogate biomarker for progression of multiple sclerosis (MS). Data from the FREEDOMS core and extension trials revealed that NEDA-4 status in the first year is a better predictor of long-term outcomes than NEDA-3. Volumetry has gained increasing attention as a possible approach to enable earlier and more accurate prognosis (1) and as a secondary outcome measure in clinical trials (2). Loss of brain tissue due to the progression of the disease may be easily overlooked during expert reading of magnetic resonance imaging (MRI). Monitoring of disease progression and therapy control may profit from quantitative volumetric approach to estimate lesion load and brain atrophy. Novel automatic methods for whole-brain volumetry (3–8) enable robust and effective longitudinal monitoring of atrophy and disease progression, while treatment goals in MS have shifted to require the setting of novel targets for disease monitoring.

Voxel-based morphometric (VBM) and surface-based morphometric group studies in MS patients (9–15) identified reduced mean overall cortical thickness and regional subcortical and cerebellar gray matter (GM) volume loss. Focal cortical thinning in frontal and temporal brain regions with a tendency to predominantly affect the left hemisphere was observed after first onset even in subgroups with mild disability. After relapses, in more severely disabled patients, cortical thinning was also evident in the pre- and postcentral gyrus. The presence of widespread cortical thinning has been suggested as a predictor for cognitive impairment, while atrophy of the superior frontal gyrus, the thalamus, and the cerebellum were independent predictors for conversion of clinically isolated syndrome (CIS) patients to MS (16). Reduced white matter (WM) volume and increased curvature of the cortical band has also been detected earlier in CIS (17). A follow-up study revealed significantly increased loss of total GM in MS patients with disability progression after 5 years (67 patients) as compared to patients with stable disability. Similar results were obtained for cortical GM and the volume of the putamen, whereas no difference was found in WM atrophy progression. At 10 years (50 patients) only a trend toward larger loss of total GM volume in patients with disability progression was detected (18).

Ahead of translating such findings into clinical practice, further research needs to address the extent and reproducibility of regional morphometric alterations in individual patients (i.e., whether they may contribute to subgroup classification and diagnostics). In this study, we suggest an automated analysis pipeline and statistical framework allowing the generation of morphometric reports that can be integrated into the diagnostic workup. Regional morphometric parameters estimated from single T1-weighted MRI datasets of individual MS patients are statistically compared to 323 healthy control (HC) datasets with a wide age range. The analysis is built on the free software packages FSL and FreeSurfer, yet the concept is not necessarily restricted to these environments. After testing the pipeline on subsets of the HC group, we provide a first proof-of-concept using 55 follow-up

MRIs in five patients with a confirmed diagnosis of relapsing–remitting MS (RR-MS) and mild to moderate disability.

## MATERIALS AND METHODS

The study was approved by the Kantonale Ethikkommission Bern. All subjects provided written informed consent and the study was performed in accordance with the Declaration of Helsinki.

### Patients

We applied our methodology to 55 consecutive MRI scans in five patients with RR-MS under treatment with natalizumab (Biogen Corp., Cambridge MA, USA). Inclusion criterion was that patients had at least ten follow-up MRI exams at our institution, covering at least 3 years of observation time. Overall, follow-up time was  $5.03 \pm 0.83$  years (mean and SD) and the time interval between scans was  $0.55 \pm 0.06$  years. Demographic patient information is provided in **Table 1**. Scanner specifications and sequences are compiled in Table S1 in Supplementary Material.

### Healthy Controls

Patient MRIs were statistically compared to a normative dataset (**Figure 1A**) consisting of HCs that were acquired during previous studies performed at the Inselspital, Bern, see Table S2 in Supplementary Material. After calculation of the region-specific morphometric parameters, only age, sex, an anonymized subject ID and MRI parameters (manufacturer, model, and sequence) were stored in the fully anonymized database as confounding variables. The data were derived from  $N = 323$  high-resolution T1-weighted (T1w) imaging datasets from 267 neurologically healthy subjects (142 datasets from male subjects and 181 from females, mean age 35.9 years, SD 18.0 years, and age range 7–79 years, see Figure S1 in Supplementary Material for the age and sex distribution). For other demographic and technical parameters, see **Table 2**. Twenty-one participants were scanned twice, two were scanned three or four times, five were scanned five times, and one was scanned six times with separation smaller than 2 years.

We used subsets of our normative dataset to explore the expected rate of statistical anomalies in HCs by testing the selected subjects against the remainder in a leave-one-out cross-validation (LOOCV). One subset encompassed 34 randomly selected MRIs from 31 HCs (10% of all HCs, but at least one MRI from each previous study) and resembled the age, sex, MR type and sequence distribution of the whole normative dataset. To test whether deviations detected in patients depended on the different age, sex, scanner, and sequence characteristics in patients and controls (see **Table 2**), we also compiled a HC subset with characteristics matched to the MRIs of patients 2 and 3.

### Lesion Detection and Filling

Multiple sclerosis lesions may affect estimates of partial volume estimates (PVE) of GM, WM, and CSF (19). Similar biases are to be expected for volume segmentations. For cortical thickness measurements, the effect can be insignificant (20). Despite the recent finding that lesion filling might occasionally yield

**TABLE 1** | Demographic and clinical information on patients.

	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5
<b>Sex</b>	Female	Female	Female	Female	Female
<b>Age (years)</b>					
At first scan	16.32	29.93	31.72	21.43	41.41
At last scan	22.19	35.24	36.78	25.09	46.67
Mean follow-up intervals (years)	0.65	0.48	0.56	0.52	0.53
Disease duration at first scan	4.33	n.a.	12.42	2.75	4.5
<b>New lesions</b>	+1 (TP2)	+1 (TP2)	+2 (TP7)	+1 (TP7)	+2 (TP2)
<b>EDSS</b>					
At first scan	1.5	n.a.	2	1.5	2
At last scan	1	n.a.	4	1	2
Annual change rate	-0.087	n.a.	0.435	-0.169	0.067
Rate different from zero	$p = 0.388$	n.a.	$p = 0.395$	$p = 0.386$	$p = 0.567$
Spearman correlation with age	$p = 0.246$	n.a.	$p < 10^{-4}$	$p = 0.036$	$p = 0.266$
<b>MR scanner</b>	Verio	Verio	Verio	Verio	Verio
<b>MR sequence</b>					
MDEFT	0	1	1	2	0
MP-RAGE van der Kouwe	10	12	10	8	11

EDSS, expanded disability status scale; MDEFT, modified-driven equilibrium Fourier transform; MP-RAGE, magnetization-prepared rapid gradient-echo; "+n (TPm)" means that n new lesions were discovered at time point m of the follow-up series, which were not present at time point m-1. Values with significance  $p < 0.05$  are displayed in bold-face.

spuriously thinned cortex near juxta-cortical lesions (15), we integrated lesion filling ahead of morphometric analysis of MS patients. To detect lesions, we first co-registered fluid-attenuated inversion recovery (FLAIR) images acquired during the same session to the T1w images. We made use of Nabla Net (21), a recently published method using a deep convolutional architecture (a fully convolutional network in which an encoder-decoder structure computes high-level features that are combined with lower-level features using skip connections) and winning algorithm at the MICCAI 2016 MSSEG challenge to automatically produce lesion masks from FLAIR images. We calculated lesion volumes at each visit and used the lesion masks as inputs to FSL's `lesion_filling` command to obtain T1w images with lesions filled by white-matter intensities. In HCs, these steps were not performed.

## Automated Processing of T1-Weighted Images

Processing of lesion filled (if applicable) T1w images was performed using the free software packages FSL and FreeSurfer on a quad-core workstation under Ubuntu Linux, release 14.04 LTS. For displaying statistics and results, self-written Octave scripts were used (CR). In the paper, two significance levels  $\alpha = 0.01$  and  $\alpha = 0.05$  are used and results are contrasted.

Details of the volumetric and morphometric analysis pipeline are provided in the Presentation S1 in Supplementary Material. In brief, the total volumes of cerebrospinal fluid (CSF), GM, and WM were estimated using FSL<sup>1</sup> [version 5.0 (22)]. For estimation of the volumes of segmentations of the GM, WM, and CSF volume we used FreeSurfer<sup>2</sup> (version 5.3.0). Regions of interest (ROIs) included segmentations of the hippocampus and the amygdala, the thalamus, the basal ganglia, the ventricles,

the corpus callosum, and the cerebellum. The procedures are described in detail in Ref. (23, 24).

For surface-based analysis (SBA), the interfaces between WM and GM as well as between GM and CSF were estimated with FreeSurfer (Figure 1B). The technical details of these procedures have been described previously (25–27). Cortical ROIs were defined by automatic parcellation of the cortex (28) according to the atlases by Desikan et al. (29) and Destrieux et al. (30) and surface-based morphometric parameters were reported as parcellation-wise averages. Details on the extraction of the following nine regional volumetric and morphometric parameters are summarized in Presentation S1 in Supplementary Material: cortical thickness (mean and SD), cortical surface area, cortical GM volume, mean and Gaussian curvature of the cortex, curvature and folding index, and the contrast between GM and WM along the cortical band.

## Quality Control

FreeSurfer's automatic surface tessellation, volume segmentation, and cortex parcellation procedures may occasionally produce errors, which in turn may decrease the accuracy of anomaly detection. This can be prevented either by labor-intensive visual quality control in each individual or by statistical rejection of outliers. To automatically reject outliers, values of raw morphometric parameters  $X$  in each ROI exceeding within-group thresholds  $X_{<}$  and  $X_{>}$  more than 1.5 interquartile ranges from the 25% and 75% percentiles of the full distribution were rejected automatically. After a provisional polynomial fit to the empirical age dependence (see Figure 1C and text below), the same procedure was repeated for the residues  $x = X - X_{\text{poly}}$ . Then, a final age fit was performed on the cleaned data. In our large control group ( $N = 323$ ), we limited artifact rejection to this automated processing (Figure 1D). In the smaller group of follow-up MRIs from the same patient ( $10 \leq N \leq 13$ , see Table 1), this rejection scheme was followed up by visual inspection of doubtful brain regions using FreeSurfer's

<sup>1</sup><http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>.

<sup>2</sup><http://surfer.nmr.mgh.harvard.edu>.

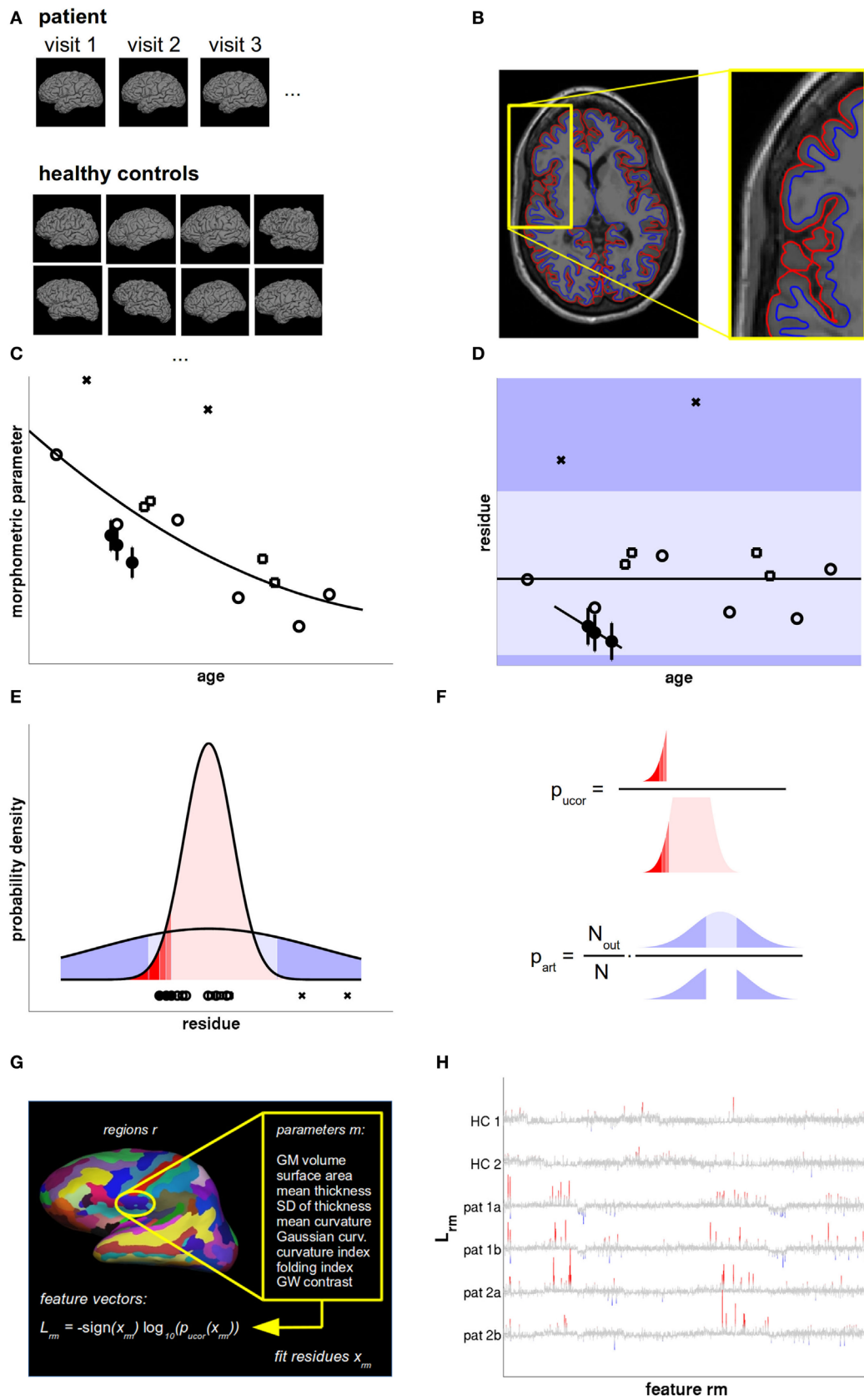


FIGURE 1 | Continued

**FIGURE 1** | Data analysis and statistical evaluation, see Section “Materials and Methods” for details. **(A)** T1w magnetic resonance imaging (MRI) acquired during all patient visits were compared with a normative database of a large number of healthy controls (HCs), accounting for age, sex, and scanner parameters. **(B)** The MRIs were processed with FreeSurfer to define the pial (red) and gray–white surface (blue). From these, the morphometric parameters of interest were calculated region-wise. **(C)** The cross-sectional age dependence of morphometric parameters in HCs (open symbols) was fitted by a low-order polynomial. The overall measurement accuracy was estimated from HCs with follow-up MRIs within short temporal separation (open squares) and displayed as error bars in patient data (filled circles). **(D)** From the fit residues outliers were rejected (crosses) and linear age trends were fitted to the patient data. The dark shaded areas symbolize the regions where artifacts could be detected as outliers. Artifacts producing small values (light shaded region) remained undetected. **(E)** Valid measurements (red) and artifacts (blue) were both modeled as Gaussians with different widths. **(F)** The uncorrected  $p$ -value  $p_{uncor}$  of an observation in patients was calculated from the distribution of fit residues in HCs. The artifact probability  $p_{art}$  was estimated from a scaled empirical outlier fraction. The scaling factor was determined from the fraction of artifacts detectable as outliers. **(G)** Feature vectors were generated as the signed base-10 logarithm of 2,976  $p$  values calculated per cortical region and morphometric parameter. **(H)** Feature vectors of two HCs and two follow-up magnetic resonance imagings (MRIs) of two multiple sclerosis (MS) patients as a function of feature number. Insignificant features ( $p > 0.01$  uncorrected) are displayed in gray. Increased values are displayed in red and decreased values in blue, with full color indicating that deviations are significant after false discovery rate (FDR) correction.

**TABLE 2** | Demographic and technical parameters of the full control dataset, random, and patient-matched subsets used for leave-one-out cross-validation (LOOCV) and the multiple sclerosis (MS) patients.

	All healthy controls (HCs)	Randomly selected HCs	Patient-matched HCs	RR-MS patients
<b>Subject number</b>	267	31	18	5
<b>Dataset number</b>	323	34	19	55
<b>Age</b>				
Mean (SD)	35.9 (18.0)	35.6 (17.8)	27.8 (5.4)	30.9 (8.7)
Range	7–79	8–72	21–39	16–46
<b>Follow-up time (years)</b>				
Mean (SD)	n.a.	n.a.	n.a.	5.03 (0.82)
Interval, mean (SD)	n.a.	n.a.	n.a.	0.55 (0.06)
<b>Sex</b>				
Male	141	13	0	0
Female	182	21	19	55
<b>MR scanner</b>				
Verio	168	17	19	53
Trio	155	17	0	2
<b>MR sequence</b>				
MDEFT	171	18	3	4
MP-RAGE standard	62	7	0	0
MP-RAGE van der Kouwe	35	3	16	51
MP-RAGE ADNI	55	6	0	0

The age and sex distribution of the control dataset is shown in Figure S1 in Supplementary Material.

ADNI, Alzheimer’s disease neuroimaging initiative; MDEFT, modified driven equilibrium Fourier transform; MP-RAGE, magnetization-prepared rapid gradient-echo; RR-MS, relapsing–remitting multiple sclerosis.

freeview software with the pial and gray–white surfaces overlaid over the T1w MRI. In rare occasions, measurements caused by artifacts were subsequently excluded manually.

After exclusion of statistical outliers from the HC dataset, the expected region- and parameter-specific measurement accuracies were estimated as the square root of the mean intra-subject squared error of the 87 MRI datasets from the  $N_{rep} = 31$  healthy subjects who had undergone more than one MRI acquisition with separation smaller than 2 years, see Presentation S1 in Supplementary Material for details.

## Standardized Result Presentation

To support expert inspection of T1w MRIs and provide additional quantitative information, we displayed the patients’

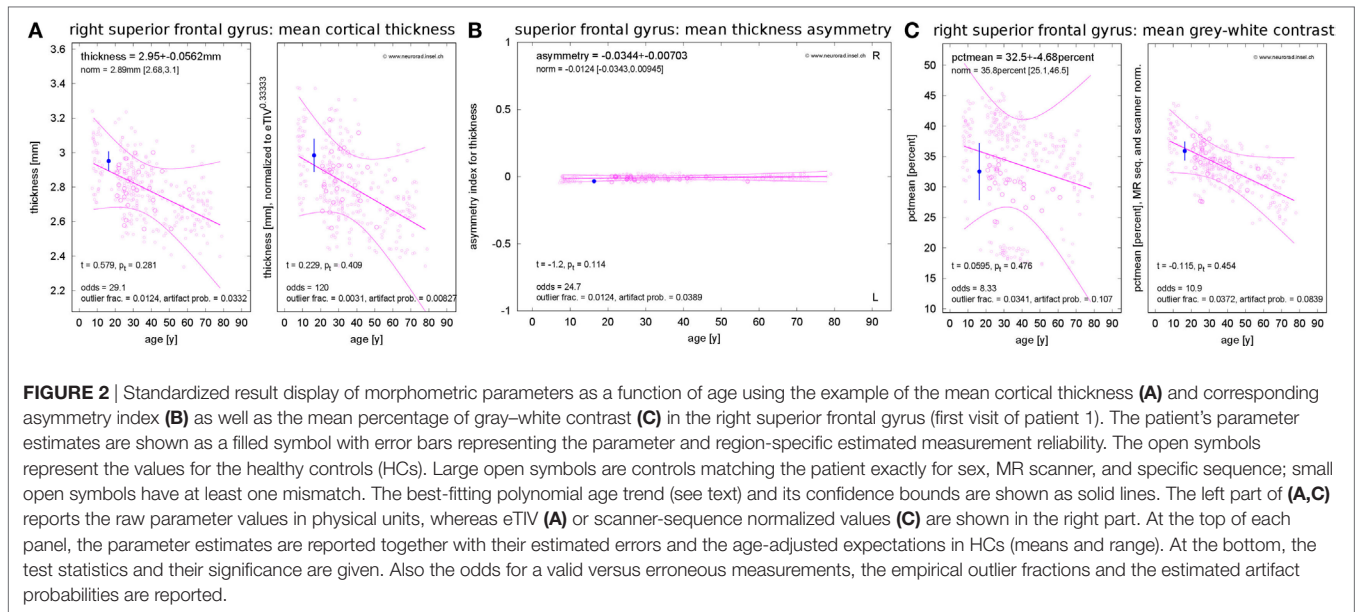
morphometric parameters and their measurement accuracies in each ROI together with the normative values as a function of age, see Figure 2. We used an in-house written Octave script (CR) to generate a standardized result display for all morphometric parameters, all volume segmentations and all cortical parcellations without preselection. All results were stored as png figures and automatically generated html pages were used to allow unrestricted navigation and switching between region-based and measure-based result compilations.

Besides the raw morphometric parameters in physical units as estimated by FreeSurfer and FSL (left part of Figure 2A), brain volume normalized versions were evaluated (right part of Figure 2A). To this end, all parameters were scaled to the mean estimated total intracranial volume (eTIV) of the full control dataset as reported by FreeSurfer. The geometrically expected scaling exponents (31) were accounted for, i.e., volumes were scaled as  $eTIV^n$  with exponent  $n = 1$ , areas with exponent  $n = 2/3$  and thicknesses as well as radii with exponent  $n = 1/3$ . According to their definition as functions of inverse radii, the mean curvature and the intrinsic curvature index were scaled with exponent  $n = -1/3$  and the Gaussian curvature and the folding index were scaled with  $n = -2/3$ .

To assess inter-hemispheric asymmetries of morphometric parameters, we estimated asymmetry indices by calculating the difference between corresponding right and left structures and normalizing to their sum (Figure 2B). For complete symmetry, this index is zero and ranges between  $-1$  for extremely left-dominated and  $+1$  for extremely right-dominated structures.

The gray–white contrast is highly dependent on the acquisition sequence and the MR scanner type. For example, the MDEFT sequence produces higher contrast than MP-RAGE sequences. This can be seen in the left part of Figure 2C, where distinct point clouds of HC data are apparent, each one stemming from a different scanner–sequence combination. To compensate this influence, we normalized contrasts to the same mean value for each scanner–sequence combination rather than to mean eTIV (right part of Figure 2C).

Since morphometric parameters, and thus the significances of observations, are age-dependent (32–38), all results were displayed as a function of age. Polynomial age trends were fitted to the estimates of all  $N = 323$  members of the normative dataset and statistical analysis was performed on the fit residues (Figures 1C–E). To avoid overfitting, the polynomial degree  $d$  was increased



stepwise between zero (constant fit) and a maximal value of  $d_{max}$  until the higher degree ceased to reduce the variance of the residues significantly ( $F$ -test for nested models). We heuristically chose  $d_{max} = N/20 = 16$  as an upper limit. A Bonferroni correction of the significance level  $\alpha$  by a factor  $1 \leq d \leq d_{max}$  was applied to account for the number of multiple tests actually required on step  $d$  of this procedure in the simplest possible way. The polynomial age trends and corresponding confidence intervals for the full cloud of the normative dataset were displayed together with the patients’ morphometric parameters; see Figure 2.

### Empirical Outlier Rate, Artifact Probabilities, and Odds for Valid vs. Erroneous Measurements

Despite outlier and artifact rejection (see above), the measurement of morphometric parameters may result in spurious values caused by image artifacts or too-weak gray–white contrast. We modeled residues  $x$  caused by artifacts by Gaussian probability densities centered at fit residue  $x = 0$ . The width  $\sigma_{out}$  was estimated by the mean absolute fit residues of all those HCs that were discarded as outliers (Figure 1E). Under these circumstances, the probability of finding artifact-related measurements as extreme as or even more extreme than the empirically detected outlier thresholds  $x_{<}$  and  $x_{>}$ , i.e., the probability  $0 \leq p_{out} \leq 1$  of detecting artifacts as outliers, is given in terms of the cumulative density function of the normal distribution with width  $\sigma_{out}$ :

$$p_{out} = \Phi_{\sigma_{out}}(x_{<}) + 1 - \Phi_{\sigma_{out}}(x_{>}). \quad (1)$$

As artifacts may occur at any magnitude, some may result in parameter estimates within the range  $[x_{<}, x_{>}]$  and these cases remain undetected by our quality control procedures. To estimate the region- and parameter-specific artifact probability, we counted the empirical outlier numbers  $N_{out}$  found in the  $N = 323$

HCs (see above) and scaled the outlier rate up by a factor  $1/p_{out}$  (see Figure 1F):

$$p_{art} = N_{out} / N * 1 / p_{out}. \quad (2)$$

We set  $p_{art} = 0$  if for a morphometric parameter or ROI no outliers were observed in HCs, taking into account that this value is only a lower limit of the true artifact probability. From the artifact probability, we calculated the region- and parameter-specific odds for a valid as opposed to an erroneous (i.e., artifact-corrupted) observation by

$$odds = (1 - p_{art}) / p_{art} \quad (3)$$

and trusted the observation when  $odds \gg 1$ .

### Statistical Assessment

Sex (38, 39), MR type (38, 40), and sequence (41) can bias morphometric estimates. To reduce confounding effects, the statistical significance of deviations between the patient and the controls was assessed by restricting the analysis to that subset of HCs who fully matched the patient’s characteristics with respect to sex, MR type, and acquisition sequence. The significance of an observed deviation of a patient’s fit residue  $x$  from the normative data can be influenced by three factors, (i) the probability of the observation being valid (as opposed to erroneous, i.e., caused by any kind of artifact), (ii) the width of the distribution of the normative data, and (iii) the measurement error of the observation itself. Treatment of issue (i) has already been discussed in the last paragraph. The larger the odds for valid measurement, the more trustworthy an observation. Issues (ii) and (iii) were treated by calculating the probability of finding an empirical observation at  $x$  if its true value  $x'$  is uncertain with a measurement error  $\sigma_{meas}$ . Assuming normally distributed fit residues in the subset of HCs used for statistics (mean at  $x = x_{norm}$  and SD  $\sigma_{norm}$ ), straightforward algebra yields that the variances add up  $\sigma^2 = \sigma_{norm}^2 + \sigma_{meas}^2$  and,

thus, the  $p$  values are given by the cumulative density function  $\Phi_\sigma$  of a Gaussian distribution with mean  $x_{\text{norm}}$  and SD  $\sigma$  in the following way (Figure 1F):

$$p_{\text{ucor}}(x) = 1 - \Phi_\sigma(x - x_{\text{norm}}) \text{ if } x > x_{\text{norm}}, \quad (4a)$$

$$p_{\text{ucor}}(x) = \Phi_\sigma(x - x_{\text{norm}}) \text{ if } x < x_{\text{norm}}. \quad (4b)$$

## Feature Vectors

To assess the reproducibility of direction and magnitude of deviations  $x_{\text{rm}}$  in ROIs  $r$  and morphometric parameters  $m$ , we defined feature vectors by signed base-10 logarithms of  $p$  values (Figure 1G)

$$L_{\text{rm}} = -\text{sign}(x_{\text{rm}}) \cdot \log_{10}(p(x_{\text{rm}})). \quad (5)$$

This yielded large values where the deviations were significant and small ones otherwise (Figure 1H). The overall length of the feature vectors was 2,976, see Table S3 in Supplementary Material.

## Estimation of Regional Atrophy Progression from Follow-up MRIs

The evaluation concept presented above can be extended to group studies (independent measurements) and follow-up analysis of individual patients (dependent measurements). If the results of more than two exams are available, one can test not only for different mean position of the fit residues  $x$  in patients and HCs but also for different age dependence. The first question can be answered by a two-sample  $t$ -test (variant for unequal group sizes and unequal variances). To address the age dependence, we fitted a linear model  $x(t) = a \cdot t + x_0$  to the patients' residues  $x$  at time points  $t$ . If this model described the patients' age trend well (i.e., the residual variance  $S^2$  was not larger than the measurement error  $\sigma_{\text{meas}}^2$  in a chi-squared test on the variance), we performed a  $z$ -test to evaluate whether the slope  $a$  differed from zero. For  $n$  measurements, the slope uncertainty is given in terms of  $S^2$  and the variance  $\sigma_t^2$  of the measurement time points by

$$\Delta a = \frac{S}{\sigma_t} \cdot \frac{1}{\sqrt{n-2}}. \quad (6)$$

We used the same approach to assess the stability of the lesion volume or of the disability level (EDSS) during the follow-up period.

## Significance Highlighting

When generating the standardized result representation, deviations of the single patient estimate or of the longitudinal change during the follow-up period from the expectation were automatically highlighted with a yellow figure background if the  $p$  value was smaller than the significance level  $\alpha$ . The highlighting indicated statistical anomalies that required secondary inspection by a trained expert. Correction for multiple comparisons during extensive testing ( $>10^3$  tests in each dataset) was performed using the concept of false discovery rate (FDR) (42). Deviations that remained significant after FDR correction were highlighted with a red figure background.

## RESULTS

### Technical Parameters

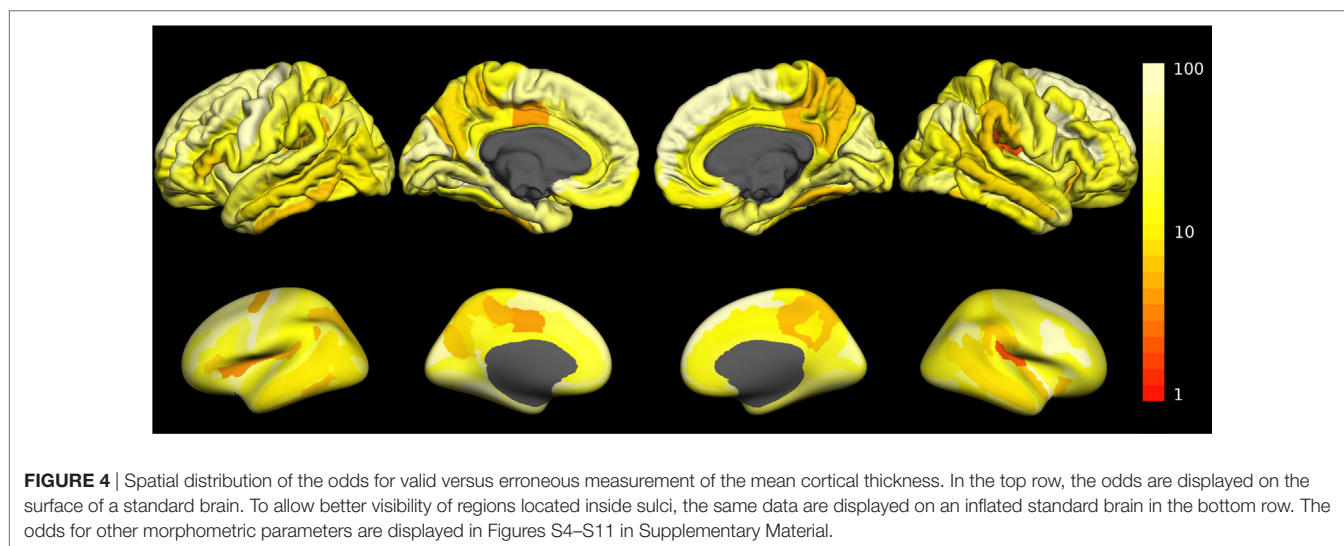
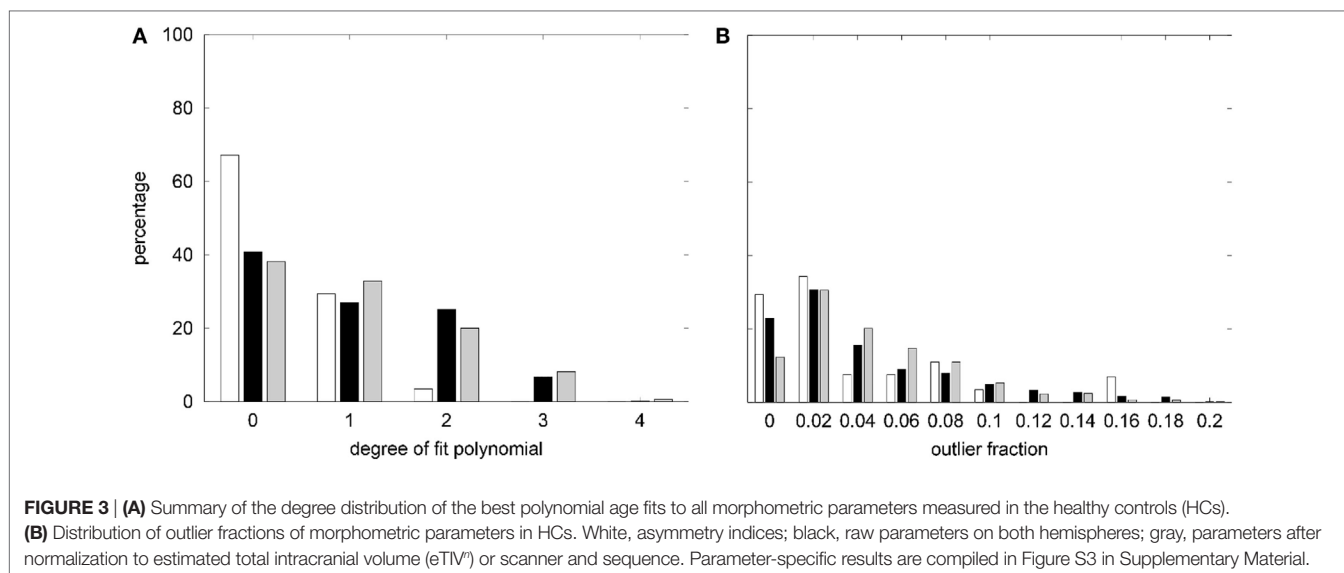
Several morphometric parameters evaluated in cortical parcellations (cortical GM volume, cortical surface area, mean and SD of the cortical thickness, four curvature measures, and the gray-white contrast) turned out to be dependent on each other in the HC data. Figure S2 in Supplementary Material reveals a strong correlation of the cortical GM volume with mean thickness and the surface area, but these measures were only weakly correlated with one another. Similarly, the SD of the cortical thickness was correlated with the mean thickness and the GM volume. The four curvature parameters were highly correlated with each other and the gray-white contrast was weakly negatively correlated with them.

The degree distribution of best low-order polynomial fits to the age trends of asymmetries, raw, and eTIV or scanner-sequence normalized morphometric parameters in the normative dataset are displayed in Figure 3A. Despite allowing a maximal degree  $d_{\text{max}} = 16$ , constant ( $d = 0$ ) and linear ( $d = 1$ ) fits were sufficient in over 90% of asymmetry indices. For morphometric parameters on the left and right hemisphere, the same was true when quadratic fits were included. The maximal polynomial degree  $d = 4$  was used in very rare cases and  $d = 3$  was necessary in a higher fraction of cortical parcellations for the mean and the SD of cortical thickness than for GM volumes, cortical surface areas, and curvature measures (see Figure S3A in Supplementary Material).

The fraction of outliers found in the normative dataset is displayed in Figure 3B and a detailed compilation is given in Figure S3B in Supplementary Material. The largest mean outlier fractions were observed for curvature measures. The spatial distribution of odds for valid versus erroneous measurements is shown in Figure 4 for the example of the mean cortical thickness. Values larger than 10 were very common and the highest odds (i.e., highest confidence in valid measurements) were located on the frontal lobes. The lowest odds (i.e., largest likelihood for artifacts in cortical thickness) were found in the cingulate gyrus and the insula. Figures S4-S11 in Supplementary Material show the corresponding information for all other cortical morphometric parameters.

### Reproducibility of Feature Vectors

The feature vectors  $L_{\text{rm}}$  encode deviations from the age- and sex-adjusted expectation in HCs. Figure 5 shows Pearson's correlation matrix between the  $L_{\text{rm}}$  calculated from different MRIs. In the HCs of the randomly selected LOOCV test set (Figure 5A), the correlation coefficients were predominantly small. Matrix elements connecting different MRI datasets of the same healthy subjects were among the largest ones (datasets 2 and 17, 3 and 8, as well as 6 and 16). By contrast, in the patient dataset (Figure 5B), a prominent block pattern became apparent, clearly reflecting patient-individual properties of the  $L_{\text{rm}}$ . The observation that correlation between feature vectors of different patients was in general higher than between different HCs might be associated with disease-related effects in RR-MS that need further investigation in a larger cohort.



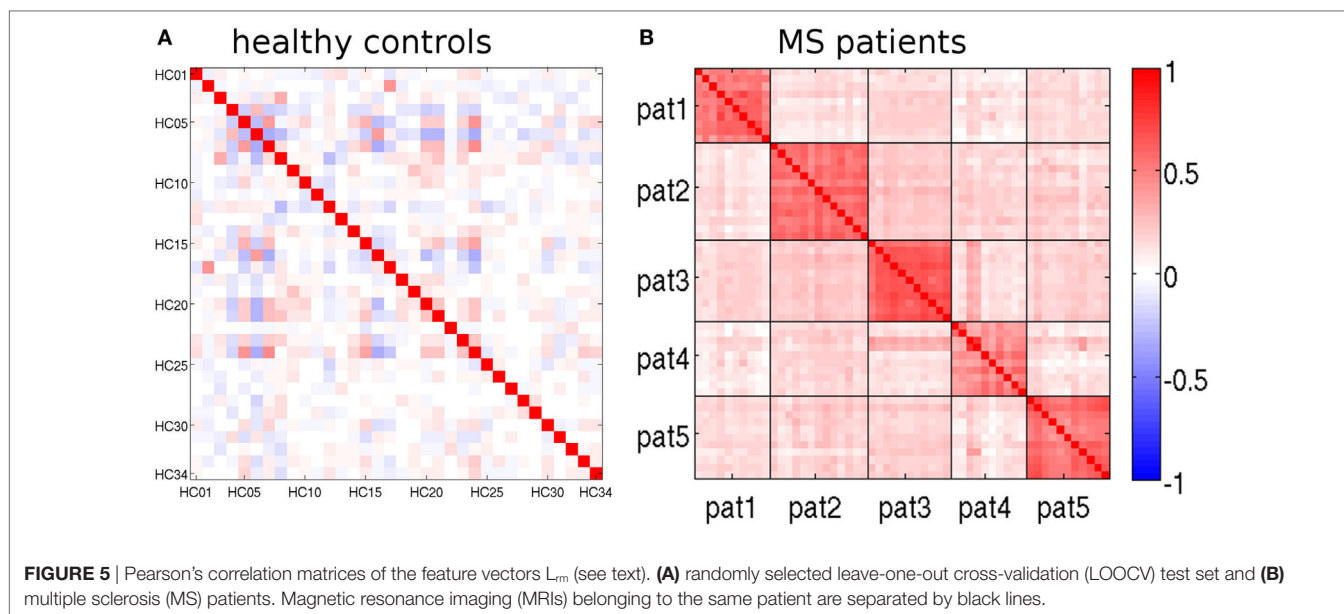
## Anomaly Detection in HCs

The overall rates of statistical anomalies with  $p < \alpha$  in the random and patient-matched LOOCV subsets and the MRIs of patients 2 and 3 are reported in **Table 3** for analysis at significance levels  $\alpha = 0.01$  and  $\alpha = 0.05$ . We used binomial tests to check the null hypotheses that the anomaly rates were given by  $\alpha$ , on the one hand, and by the empirical rate obtained in the random LOOCV subset on the other. At significance level  $\alpha = 0.01$ , anomaly rates without FDR correction (yellow background in our standard result representation) were between 1.50 and 1.74% in HCs. This turned out significantly larger than expected (Bonferroni corrected for multiple comparisons). With the exception of the asymmetry indices this was not the case when repeating the analysis at a significance level  $\alpha = 0.05$ . The overall anomaly rates in the patient-matched HCs (uncorrected for multiple comparisons) were not different from the ones empirically found

in the randomly selected HCs. However, FDR-corrected anomaly detections (red background in our standard result representation) were smaller in the patient-matched HCs than in the random HC subset.

When comparing patient MRIs to LOOCV in HCs, the anomaly rates were elevated by at least a factor of 2. All differences were highly significant with  $p = 0$  to machine precision. Anomaly rates for all morphometric parameters separately are compiled in **Table S4** in Supplementary Material. With the exception of the PVE (and to a lesser extent also of the SDs of the cortical thickness) anomaly rates were elevated in the patients for all morphometric parameters. This effect was less prominent when anomaly detection was FDR-corrected to account for multiple comparisons. In healthy subjects, the rate of (potentially false) detections was higher for the Gaussian curvature, the curvature and the folding index than for the other morphometric parameters.





## Exemplary Use Case: Progressive Volume Loss in Patients with Relapsing–Remitting Multiple Sclerosis

A study evaluating the usefulness of the proposed analysis pipeline for single time-point MRIs in a larger number of temporal lobe epilepsy patients is presented elsewhere (43). Here, we concentrate on progressive volumetric and morphometric changes detectable in individual patients with RR-MS. During observation time all patients had only a small amount of new lesions (Table 1) and none of the slopes of linear fits to the lesion volumes estimated from the FLAIR images was statistically different from zero (Table 4). The slope of a linear fit to the EDSS disability index was not different from zero, either (Table 1). However, the Spearman correlation coefficient of EDSS with age was significant for patients 3 and 4.

Table 4 summarizes results for the PVE of all patients. With the exception of patient 4, none of the PVEs deviated significantly from the age-adjusted expectation (significance level  $\alpha = 0.05$ ), neither at the first nor at the last MRI exam of the observation period. In patient 4, we detected reduced GM volume for a female of 21 years at the first MRI exam ( $p = 0.008$ ). At the last exam, brain volume was too small and CSF volume too large (both  $p = 0.034$ ) for a female of 25 years. PVE changes were often significantly larger than expected from cross-sectional aging in the control group. Brain volume loss and CSF volume increase were stronger than expected in 4/5 patients and GM loss was stronger than expected in 3/5 patients. Examples of GM volume loss and CSF volume increase in patient 1 are presented in Figure 6. Indeed, the volumes were still within the normal ranges, but progressive decline over the past 6 years deviated from the expectation ( $p = 0.003$  for GM and  $p < 10^{-3}$  for CSF, both after normalization for eTIV).

The fraction of patients presenting with deviations of volume segmentations, of mean cortical thickness, or of the gray–white contrast deviations of from age- and sex-adjusted expectations

in the first and last MRI is compiled in Figures S12–S17 in Supplementary Material. Here, we focus on more severe than expected linear changes of these parameters. For the volumes of the lateral ventricles, the brain stem, and the right-hemispheric cerebellar GM, we observed deviations from the expected change rate in both directions, see Figure 7. Note that the blue-to-white color scale can either indicate significantly stronger than expected decrease or weaker than expected increase and vice versa for the red-to-yellow colorbar. An example of more severe than expected volume loss of the right caudate volume (found in 4/5 patients, same for the left hemisphere) in patient 3 is given in Figure 8 ( $p = 0.002$  after normalization for eTIV).

Concerning changes of mean cortical thickness in individual patients, we found deviations from the expectation in both directions, see Figure 9. Weaker than expected thickness decrease (or even increase) was observed predominantly in the bilateral temporal lobes and insula. By contrast, stronger than expected thickness decrease was observed predominantly in the bilateral cingulate, parieto-occipital regions and in the pre- and postcentral gyri and sulci. In Figure 10, we demonstrate at the example of the short insular gyri on the left that the coexistence of thickness increase and decrease is not an artifact of our methodology. We found pronounced thickness increase in patient 1 (Figure 10A), whereas marked thickness decrease was detected in patient 5 (Figure 10B).

## DISCUSSION

### Methodological Developments

Several studies using MR-based morphometry have indicated that GM atrophy in RR-MS emerges not only as a global but also rather regional process, affecting the temporal lobe, the pre- and postcentral gyrus, cingulate gyrus, basal ganglia, and the thalamus (44, 45). Long-term atrophy progression of GM was found

**TABLE 3** | Overall assessment of anomalies found in the randomly selected and the patient-matched LOOCV test sets as well as in the patient datasets at two significance levels  $\alpha = 0.01$  and  $\alpha = 0.05$ .

		Test count	$p < 0.01$		$p < 0.05$	
			$p_{\text{Uncorr}}$	$p_{\text{FDR}}$	$p_{\text{Uncorr}}$	$p_{\text{FDR}}$
<b>LOOCV: 34 random healthy controls (HCs)</b>						
Raw	Count	67,558	1,088	343	3331	464
	Percentage		1.61%	0.51%	4.93%	0.69%
	p_bino (nominal)		<b>0</b>	n.a.	0.207	n.a.
	p_bino (empirical)		–	–	–	–
Estimated total intracranial volume (eTIV) and scanner/sequence normalized	Count	67,558	1,116	349	3548	462
	Percentage		1.65%	0.52%	5.25%	0.68%
	p_bino (nominal)		<b>0</b>	n.a.	0.001	n.a.
	p_bino (empirical)		–	–	–	–
Asymmetry	Count	33,524	582	56	2,244	114
	Percentage		1.74%	0.17%	6.69%	0.34%
	p_bino (nominal)		<b>0</b>	n.a.	<b>0</b>	n.a.
	p_bino (empirical)		–	–	–	–
<b>LOOCV: 19 matched HCs</b>						
Raw	Count	37,753	567	126	1985	164
	Percentage		1.50%	0.33%	5.26%	0.43%
	p_bino (nominal)		<b>0</b>	n.a.	0.012	n.a.
	p_bino (empirical)		0.048	<b><i>&lt;10<sup>-6</sup></i></b>	0,002	<b><i>&lt;10<sup>-9</sup></i></b>
eTIV and scanner/sequence normalized	Count	37,753	578	129	2,030	170
	Percentage		1.53%	0.34%	5.38%	0.45%
	p_bino (nominal)		<b>0</b>	n.a.	<i>&lt;10<sup>-3</sup></i>	n.a.
	p_bino (empirical)		0.033	<b><i>&lt;10<sup>-6</sup></i></b>	0,140	<b><i>&lt;10<sup>-8</sup></i></b>
Asymmetry	Count	18,734	279	17	1,274	25
	Percentage		1.49%	0.09%	6.80%	0.13%
	p_bino (nominal)		<b><i>&lt;10<sup>-9</sup></i></b>	n.a.	<b>0</b>	n.a.
	p_bino (empirical)		0.005	0.004	0.283	<b><i>&lt;10<sup>-7</sup></i></b>
<b>Multiple sclerosis (MS) patients 2 and 3: 19 magnetic resonance imagings (MRIs)</b>						
Raw	Count	37,753	1460	528	3,533	736
	Percentage		3.87%	1.40%	9.36%	1.95%
	p_bino (nominal)		<b>0</b>	n.a.	<b>0</b>	n.a.
	p_bino (empirical)		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
eTIV and scanner/sequence normalized	count	37,753	1,530	545	3,736	764
	percentage		4.05%	1.44%	9.90%	2.02%
	p_bino (nominal)		<b>0</b>	n.a.	<b>0</b>	n.a.
	p_bino (empirical)		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
Asymmetry	Count	18,734	879	183	2,256	329
	Percentage		4.69%	0.98%	12.04%	1.76%
	p_bino (nominal)		<b>0</b>	n.a.	<b>0</b>	n.a.
	p_bino (empirical)		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

The randomly selected subset (34 MRI datasets) was tested against the expected error rate  $\alpha$  (binomial tests). By contrast, the MS patient and patient-matched subsets were also tested against the empirical anomaly rates obtained in the randomly selected subset. *p* Values in bold-face indicate significantly enlarged and in italic significantly decreased values (after FDR correction for multiple comparisons). Specific results of morphometric parameters are compiled in Table S4 in Supplementary Material. FDR, false discovery rate; LOOCV, leave-one-out cross-validation; n.a., not applicable.

associated with disability progression, whereas no effect was found for WM (18). To monitor longitudinal GM changes, we have employed a fully automatic pipeline for volumetric and morphometric evaluation of individual and follow-up high-resolution T1w MRI datasets (**Figure 1**) based on the free software packages FSL and FreeSurfer. We referenced our method to a large and extendable control group of currently  $N = 323$  healthy subjects exhibiting a population-based distribution and tailored statistical post-processing accounting for age, sex, brain size, gray-white contrast, and technical parameters as potential confounders. The reliability of regional estimates of morphometric parameters was

estimated from repeated MRI acquisitions with separation less than 2 years in a subset of 31 of our HCs (87 MRIs). Similarly, the HC outlier fraction observed in specific morphometric parameters and brain regions was used to estimate the corresponding artifact probability and, finally, the odds for valid versus erroneous observations. The larger these odds, the more the software's estimates can be trusted.

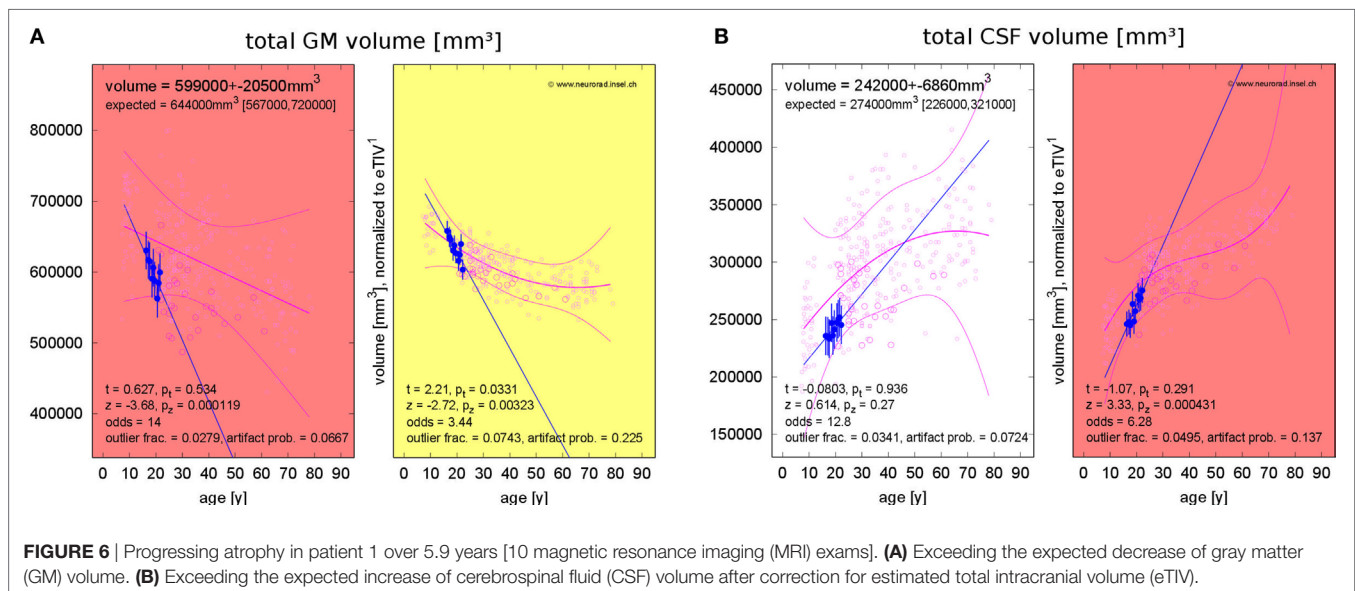
All these quantifiers were integrated into a standardized result presentation (**Figure 2**) and can be made available to the expert to provide quantitative support for clinical MRI review. As some of the surface-based morphometric parameters (especially the

**TABLE 4** | Partial volume estimates (PVE) at first and last magnetic resonance imaging (MRI) exam and annual rate of volume change.

	Pat. 1		Pat. 2		Pat. 3		Pat. 4		Pat. 5	
	(ml)	p-Value	(ml)	p-Value	(ml)	p-Value	(ml)	p-Value	(ml)	p-Value
<b>GM volume (ml)</b>										
First MRI	630	0.08	514	-0.494	514	0.462	<b>565</b>	<b>-0.008</b>	588	0.073
Last MRI	538	-0.284	503	0.485	511	0.463	527	-0.088	574	0.077
Annual change rate	<b>-11.5</b>	<b>-0.003</b>	-5.2	-0.142	<b>-5.4</b>	<b>-0.036</b>	<b>-11.8</b>	<b>-0.032</b>	-2.9	-0.209
<b>WM volume (ml)</b>										
First MRI	492	-0.207	446	-0.256	463	-0.484	496	0.213	499	-0.231
Last MRI	480	0.483	438	-0.206	464	-0.448	473	-0.294	496	-0.262
Annual change rate	-2.8	0.136	-1	0.434	<b>2.1</b>	<b>0.046</b>	3.4	0.057	0	0.296
<b>CSF volume (ml)</b>										
First MRI	236	-0.286	259	0.19	245	-0.49	289	0.174	283	0.481
Last MRI	245	0.429	269	0.101	257	0.377	<b>285</b>	<b>0.034</b>	284	0.468
Annual change rate	<b>2.8</b>	<b>4 × 10<sup>-4</sup></b>	<b>2.5</b>	<b>0.043</b>	<b>0.9</b>	<b>0.025</b>	-1.4	-0.302	<b>1.4</b>	<b>0.047</b>
<b>Brain volume (ml)</b>										
First MRI	1,120	0.286	959	-0.19	977	0.49	1060	-0.174	1,090	-0.481
Last MRI	1,020	-0.429	941	-0.101	975	-0.377	<b>1,000</b>	<b>-0.034</b>	1,070	-0.468
Annual change rate	<b>-14.2</b>	<b>-4 × 10<sup>-4</sup></b>	<b>-6.3</b>	<b>-0.043</b>	<b>-3.3</b>	<b>-0.025</b>	-8.4	0.302	<b>-2.9</b>	<b>-0.047</b>
<b>Lesion volume (ml)</b>										
First MRI	7.08	n.a.	11.3	n.a.	13.12	n.a.	8.39	n.a.	4.04	n.a.
Last MRI	5.74	n.a.	11.31	n.a.	14.28	n.a.	8.42	n.a.	3.19	n.a.
Annual change rate	-0.05	-0.388	-0.03	-0.401	0.12	0.536	0.5	0.774	-0.07	-0.389

Values with significance  $p < 0.05$  after eTIV correction are displayed in bold-face. The sign of the p-value indicates whether the observed deviation was larger or smaller than expected.

CSF, cerebral-spinal fluid; GM, gray matter; WM, white matter.

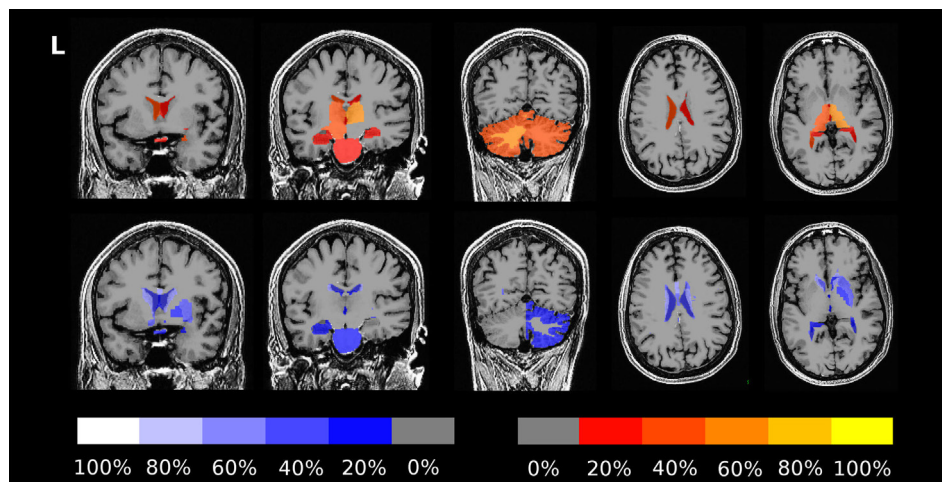


**FIGURE 6** | Progressing atrophy in patient 1 over 5.9 years [10 magnetic resonance imaging (MRI) exams]. **(A)** Exceeding the expected decrease of gray matter (GM) volume. **(B)** Exceeding the expected increase of cerebrospinal fluid (CSF) volume after correction for estimated total intracranial volume (eTIV).

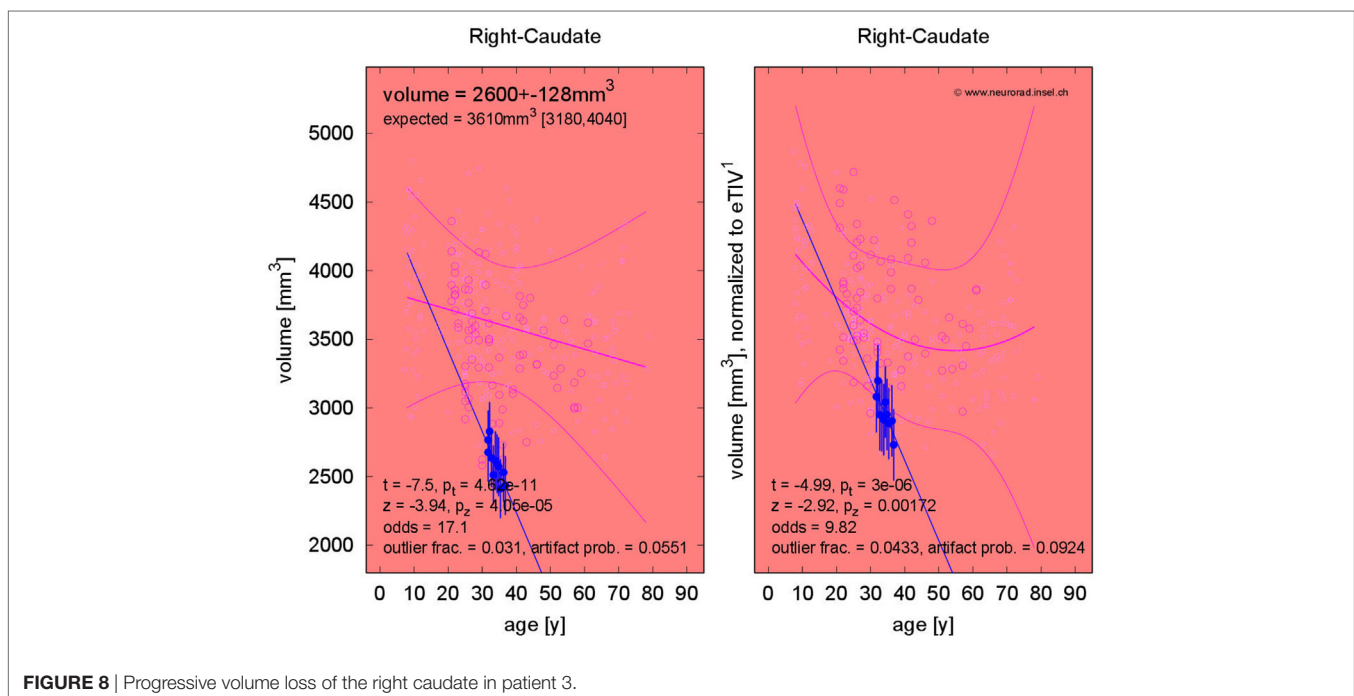
surface area and curvatures of the highly folded cortical band) are barely amenable to visual inspection of MR slices, this complements the information available to the human expert.

The analysis pipeline underwent LOOCV in HCs. In a randomly selected subset of 34 HCs, it yielded low but significantly elevated anomaly rates (see **Table 3**). These deviations were most likely due to a slight violation of the implicit normality

assumption by distributions with heavier tails. Testing a subset of 19 patient-matched HC datasets against the empirically expected anomaly rates, no major deviations were observed. By contrast, anomaly rates were significantly elevated by at least a factor of 2 in individual analysis of 19 follow-up MRI exams of two MS patients (see **Table 3** and **Table S4** in Supplementary Material).



**FIGURE 7** | Percentage of patients showing more severe than expected linear changes of volume segmentations during the follow-up period. The top row shows deviations toward stronger increase or weaker decrease on a red-to-yellow color scale. The bottom row shows deviations toward stronger decrease or weaker increase on a blue-to-white color scale. The images are in neurological orientation, i.e., the left side of the images correspond to the left hemisphere.

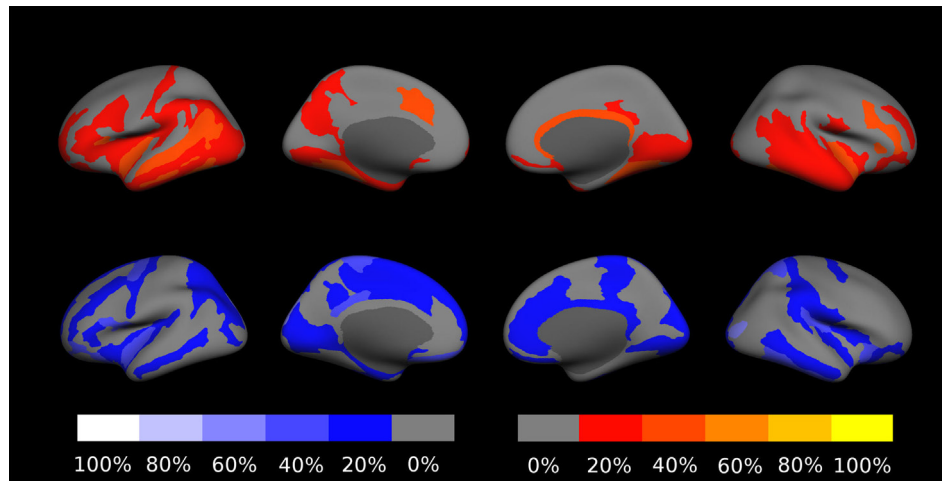


**FIGURE 8** | Progressive volume loss of the right caudate in patient 3.

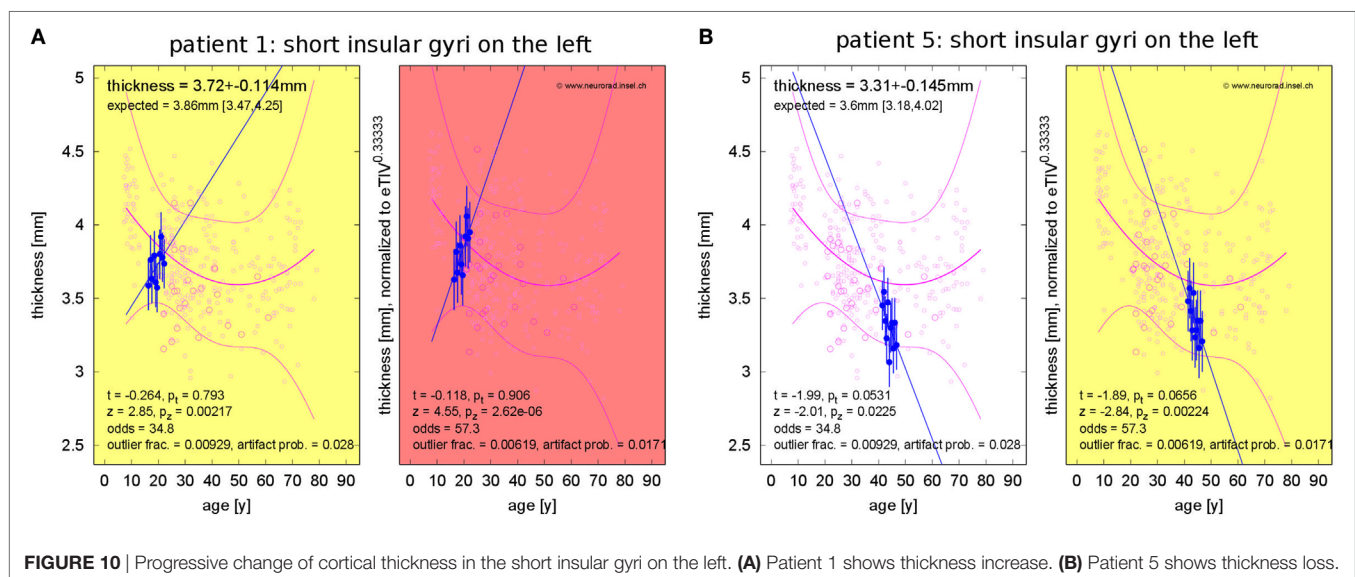
## Use Case: Detecting Regional Atrophy in RR-MS

As a proof-of-concept, we applied the analysis to one use case, i.e., the longitudinal follow-up in patients with RR-MS (5 patients, 55 MRIs with a mean follow-up time of 5 years). Despite low to moderate disability in all five patients, the expanded disability status scale (EDSS) correlated with the time elapsed during the observation period in two of the patients (Table 1). Patient 3 had an extensive duration of disease and patient 4 had abnormal PVE in the first and last MRI of the follow-up series (Table 4).

In all other patients, the PVE were normal in all MRIs of the follow-up series. Reduced cortical thickness in the first or last MRI of the follow-up series was found predominantly in the pre- and postcentral gyrus of both hemispheres and too small gray-white contrast was predominant in the left lateral temporal lobe (see Figures S14–S17 in Supplementary Material). Atrophy of deep GM structures was also observed in individual MRIs (see Figures S12 and S13 in Supplementary Material). These regional atrophy patterns are consistent with previous MS group studies (9–16) and have been assigned to cognitive impairment in MS.



**FIGURE 9** | Percentage of patients showing excess linear change of cortical thickness during the follow-up period. The top row shows deviations toward stronger than expected increase or weaker than expected decrease on a red-to-yellow color scale. The bottom row shows deviations toward stronger decrease or weaker increase on a blue-to-white color scale. In Figure S18 in Supplementary Material, a similar result compilation is shown for changes in gray-white contrast.



**FIGURE 10** | Progressive change of cortical thickness in the short insular gyri on the left. **(A)** Patient 1 shows thickness increase. **(B)** Patient 5 shows thickness loss.

According to Steenwijk et al. (45) “cortical atrophy in multiple sclerosis occurs in a non-random manner and develops (at least partly) according to distinct anatomical patterns.” Thus, analysis of regional atrophy patterns need to be further explored as markers for clinical deterioration.

In addition to analysis of morphometric anomalies in volume segmentations and cortical parcellations at a single time point, the proposed pipeline was used to detect and quantify progression of regional atrophy by fitting a linear model and classifying whether the change rate deviated from the cross-sectional expectation (Figures 6–10). After normalization to eTIV, we found stronger than expected brain volume loss in 4/5 patients (Table 4). Cortical thickness change as a regional atrophy progression marker in individuals differed from the expectation in both directions, i.e.,

stronger than expected as well as weaker than expected thickness loss or even increase (Figures 9 and 10).

## Statistical and Methodological Considerations

Morphometric parameters vary systematically with age and sex (32–39). In addition, the performance of morphometry tools depends on image quality, which in turn is influenced by technical factors, such as MR scanner type and manufacturer, field strength, and acquisition sequence (39–41). In comparison between two groups of sizes  $N_{smi}$  and  $N_{lrg}$ , these issues can be taken into account either by matching patients and HCs as exactly as possible or by introducing nuisance variables into the analysis. Unfortunately,

at the clinically relevant limit where the smaller “group” is an individual patient or MRI dataset ( $N_{\text{smi}} = 1$ ), the number of perfectly matched HCs available at most institutions is usually very small and application of nuisance parameters is mathematically hindered by the problem of inverting singular matrices.

A further difficulty with group discrimination is that statistical power crucially depends on the size of the smaller group and may, thus, be very weak in the extreme case where  $N_{\text{smi}} = 1$ . To deal with these problems, we analyzed residues after fitting low-order polynomial age trends to the morphometric parameters of the HCs (**Figure 1**) and restricted statistical testing to a suitable subset of HCs that matched the patient in all categorical characteristics (i.e., sex, MR type, and acquisition sequence). This constituted a compromise between minimization of potential sources of bias and maximizing the size  $N_{\text{reg}}$  of the accessible control group—and thus the statistical power.

Our normative dataset contained neither HCs below 7 years nor subjects older than 79 years and had a gap between 13 and 17 years. It is important to note that application of the proposed statistical analysis of fit residues to patients with ages near or beyond the borders of the controls’ age range must be considered with extreme caution. The reason is that extrapolation of the polynomial fit to the age dependence of the included HCs rapidly loses reliability outside the age range covered by the normative dataset and in consequence might introduce large systematic errors in and beyond these border regions.

Direct comparison of longitudinal age trends derived from follow-up exams in the same patients with the cross-sectional age trend in the HC group must also be approached with caution. Our normative database contains repeated MRIs only from a relatively small number of subjects. As the temporal separation between repeated MRI exams was no longer than 2 years (i.e., small as compared to the whole age range of 7–79 years), we used these datasets to estimate measurement reliabilities rather than healthy aging trajectories. For future applications, one should consider building normative databases of follow-up MRIs in HCs.

In the current version of our pipeline, we omitted FreeSurfer’s time-consuming longitudinal analysis stream. Whereas, for vertex-wise analysis, it is expected that intra-subject image coregistration would increase reproducibility and accuracy of repeated measurements; due to the implicit averaging procedures, we do not expect great improvements in our current ROI-wise approach.

Besides volumes of total CSF, GM, and WM and volumes of segmentations, our pipeline primarily uses surface-based morphometric (SBA) parameters provided by the software FreeSurfer and reports region-wise averages. SBA has advantages, both over VBM analysis as well as over expert analysis. Classic VBM is limited to regional GM volumes and concentrations. Although non-standard VBM techniques are available to estimate cortical thickness (46, 47), their test–retest reliabilities were lower than for surface-based measurement in a population of dementia patients (48). Other morphometric parameters (cortical surface area and curvature) have not yet been assessed with VBM, neither are they easily accessible to expert MRI inspection. During slice-by-slice image analysis by an expert neuroradiologist, abnormal surface areas or curvatures in the highly curved cortical band remain undetectable to the reader.

## CONCLUSION

We propose a pipeline for automated morphometric analysis of individual and follow-up MRI exams. In the present study, the analysis pipeline, the outlier and artifact handling, and the statistical post-processing were developed and the feasibility was demonstrated. In future work, the tool needs to be evaluated in larger groups of MS patients and in patients with other pathologies. Importantly, the analysis concept presented here is by no means limited to FSL and FreeSurfer results as input data. An extension to input data generated by different software packages needs to be evaluated. Extension of our current segmentation- and parcellation-wise analysis to voxel- and vertex-wise analysis must also be addressed.

A first application to 55 datasets of five MS patients enabled us to detect critical regions of stable and progressing atrophy, which matched the predilection areas identified during group studies, i.e., focal thinning and loss of gray–white contrast in frontal, central, and temporal cortex regions as well as subcortical and cerebellar GM volume loss. Further studies with larger patient groups are necessary to confirm these preliminary findings. In addition, it remains to be investigated, whether individual atrophy progression or stability in these regions may be employed as surrogate markers for the efficiency of disease-modifying treatment in RR-MS.

Application to follow-up MRIs of patients and HCs proved high correlation of the feature vectors  $L$  of Eq. 5 within the same subject as well as within the RR-MS group. We, thus, envision that the feature vectors would be promising candidates for input to machine learning algorithms to classify not only subjects but also syndromes in MS and other pathologies such as epilepsy and dementia subtypes.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Kantonale Ethikkommission Bern with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Kantonale Ethikkommission Bern (protocol 2016-02035).

## AUTHOR CONTRIBUTIONS

CR, FW, and RW designed the study. CR and RM contributed software. CR and FA analyzed the imaging data. FA, FW, AS, and AC contributed clinical information. CR, FW, and RW wrote the manuscript. CR, FA, RM, FW, AS, AC, and RW reviewed and approved the final version of the manuscript.

## ACKNOWLEDGMENTS

The authors thank Kathrin Bieri, Regula Everts, Raimund Kottke, Stefan Schweizer and Rajeev K. Verma for contributing imaging data from healthy controls for secondary use and Christian Güder

and Martin Zbinden for supporting us in distributed computing. We thank our MR technicians for their assistance during data acquisition.

## FUNDING

This study was supported by the Swiss Multiple Sclerosis Foundation. The funders had no role in study design, data

collection and analysis, decision to publish, or preparation of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/articles/10.3389/fneur.2017.00727/full#supplementary-material>.

## REFERENCES

- Geurts JGG, Calabrese M, Fisher E, Rudick RA. Measurement and clinical effect of grey matter pathology in multiple sclerosis. *Lancet Neurol* (2012) 11:1082–92. doi:10.1016/S1474-4422(12)70230-2
- Kappos L, De Stefano N, Freedman MS, Cree BA, Radue EW, Sprenger T, et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. *Mult Scler* (2016) 22:1297–305. doi:10.1177/1352458515616701
- Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* (2002) 17:479–89. doi:10.1006/nimg.2002.1040
- Durand-Dubief F, Belaroussi B, Armspach JP, Dufour M, Roggerone S, Vukusic S, et al. Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. *AJNR Am J Neuroradiol* (2012) 33:1918–24. doi:10.3174/ajnr.A3107
- Smeets D, Ribbens A, Sima DM, Cambron M, Horakova D, Jain S, et al. Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain Behav* (2016) 6:e00518. doi:10.1002/brb3.518
- Wang C, Beadnall HN, Hatton SN, Bader G, Tomic D, Silva DG, et al. Automated brain volumetrics in multiple sclerosis: a step closer to clinical application. *J Neurol Neurosurg Psychiatry* (2016) 87:754–7. doi:10.1136/jnnp-2015-312304
- Jain S, Sima DM, Ribbens A, Cambron M, Maertens A, Van Hecke W, et al. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *Neuroimage Clin* (2015) 8:367–75. doi:10.1016/j.nicl.2015.05.003
- Steenwijk MD, Amiri H, Schoonheim MM, de Sitter A, Barkhof F, Pouwels PJW, et al. Agreement of MSmetrix with established methods for measuring cross-sectional and longitudinal brain atrophy. *Neuroimage Clin* (2017) 15:843–53. doi:10.1016/j.nicl.2017.06.034
- Sailer M, Fischl B, Salat D, Tempelmann C, Schönfeld MA, Busa E, et al. Focal thinning of the cerebral cortex in multiple sclerosis. *Brain* (2003) 126:1734–44. doi:10.1093/brain/awg175
- Bodini B, Khaleeli Z, Cercignani M, Miller DH, Thompson AJ, Ciccarelli O. Exploring the relationship between white matter and gray matter damage in early primary progressive multiple sclerosis: an in vivo study with TBSS and VBM. *Hum Brain Mapp* (2009) 30:2852–61. doi:10.1002/hbm.20713
- Ramasamy DP, Benedict RHB, Cox JL, Fritz D, Abdelrahman N, Hussein S, et al. Extent of cerebellum, subcortical and cortical atrophy in patients with MS: a case-control study. *J Neurol Sci* (2009) 282:47–54. doi:10.1016/j.jns.2008.12.034
- Calabrese M, Rinaldi F, Mattisi I, Grossi P, Favaretto A, Atzori M, et al. Widespread cortical thinning characterizes patients with MS with mild cognitive impairment. *Neurology* (2010) 74:321–8. doi:10.1212/WNL.0b013e31818cbcd03
- Lansley J, Mataix-Cols D, Grau M, Radua J, Sastre-Garriga J. Localized grey matter atrophy in multiple sclerosis: a meta-analysis of voxel-based morphometry studies and associations with functional disability. *Neurosci Biobehav Rev* (2013) 37:819–30. doi:10.1016/j.neubiorev.2013.03.006
- Narayana PA, Govindarajan KA, Goel P, Datta S, Lincoln JA, Cofield SS, et al. Regional cortical thickness in relapsing remitting multiple sclerosis: a multicenter study. *Neuroimage Clin* (2013) 2:120–31. doi:10.1016/j.nicl.2012.11.009
- Tillema JM, Hulst HE, Rocca MA, Vrenken H, Steenwijk MD, Damjanovic D, et al. Regional cortical thinning in multiple sclerosis and its relation with cognitive impairment: a multicenter study. *Mult Scler* (2016) 22:901–9. doi:10.1177/1352458515607650
- Calabrese M, Rinaldi F, Mattisi I, Bernardi V, Favaretto A, Perini P, et al. The predictive value of gray matter atrophy in clinically isolated syndromes. *Neurology* (2011) 77:257–63. doi:10.1212/WNL.0b013e318220abd4
- Deppe M, Marinell J, Krämer J, Duning T, Ruck T, Simon OJ, et al. Increased cortical curvature reflects white matter atrophy in individual patients with early multiple sclerosis. *Neuroimage Clin* (2014) 6:475–87. doi:10.1016/j.nicl.2014.02.012
- Jacobsen C, Hagemeyer J, Myhr K-M, Nyland H, Lode K, Bergsland N, et al. Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study. *J Neurol Neurosurg Psychiatry* (2014) 85:1109–15. doi:10.1136/jnnp-2013-306906
- Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum Brain Mapp* (2012) 33:2062–71. doi:10.1002/hbm.21344
- Biberacher V, Schmidt P, Keshavan A, Boucard CC, Righart R, Sämann P, et al. Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. *Neuroimage* (2016) 142:188–97. doi:10.1016/j.neuroimage.2016.07.035
- McKinley R, Wepfer R, Gundersen T, Wagner F, Chan A, Wiest R, et al. Nabla-net: a deep dag-like convolutional architecture for biomedical image segmentation. In: Crimi A, Menze B, Maier O, Reyes M, Winzick S, Handels H, editors. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer (2016).
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* (2004) 23(Suppl 1):S208–19. doi:10.1016/j.neuroimage.2004.07.051
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* (2002) 33:341–55. doi:10.1016/S0896-6273(02)00569-X
- Fischl B, Salat DH, van der Kouwe AJ, Makris N, Ségonne F, Quinn BT, et al. Sequence-independent segmentation of magnetic resonance images. *Neuroimage* (2004) 23(Suppl 1):S69–84. doi:10.1016/j.neuroimage.2004.07.016
- Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* (1999) 9:179. doi:10.1006/nimg.1998.0395
- Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* (1999) 9:195. doi:10.1006/nimg.1998.0396
- Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* (1999) 8:272–84.
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, et al. Automatically parcellating the human cerebral cortex. *Cereb Cortex* (2004) 14:11–22. doi:10.1093/cercor/bhg087
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* (2006) 31:968–80. doi:10.1016/j.neuroimage.2006.01.021
- Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* (2010) 53:1–15. doi:10.1016/j.neuroimage.2010.06.010
- Im K, Lee JM, Lyttelton O, Kim SH, Evans AC, Kim SI. Brain size and cortical thickness in the adult human brain. *Cereb Cortex* (2008) 18:2181–91. doi:10.1093/cercor/bhm244

32. Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* (2001) 14:21–36. doi:10.1006/nimg.2001.0786
33. Smith CD, Chebrolu H, Wekstein DR, Schmitt FA, Markesbery WR. Age and gender effects on human brain anatomy: a voxel-based morphometric study in healthy elderly. *Neurobiol Aging* (2007) 28:1075–87. doi:10.1016/j.neurobiolaging.2006.05.018
34. Draganski B, Ashburner J, Hutton C, Kherif F, Frackowiak RS, Helms G, et al. Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *Neuroimage* (2011) 55:1423–34. doi:10.1016/j.neuroimage.2011.01.052
35. Becker JT, Duara R, Lee CW, Teverovsky L, Snitz BE, Chang CC, et al. Cross-validation of brain structural biomarkers and cognitive aging in a community-based study. *Int Psychogeriatr* (2012) 24:1065–75. doi:10.1017/S1041610212000191
36. Wierenga L, Langen M, Ambrosino S, van Dijk S, Oranje B, Durston S. Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. *Neuroimage* (2004) 96:67–72. doi:10.1016/j.neuroimage.2014.03.072
37. Wierenga LM, Langen M, Oranje B, Durston S. Unique developmental trajectories of cortical thickness and surface area. *Neuroimage* (2014) 87:120–6. doi:10.1016/j.neuroimage.2013.11.010
38. Potvin O, Dieumegarde L, Duchesne S; Alzheimer's Disease Neuroimaging Initiative. Normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain. *Neuroimage* (2017) 156:315–39. doi:10.1016/j.neuroimage.2017.05.019
39. Murphy DG, DeCarli C, McIntosh AR, Daly E, Mentis MJ, Pietrini P, et al. Sex differences in human brain morphometry and metabolism: an in vivo quantitative magnetic resonance imaging and positron emission tomography study on the effect of aging. *Arch Gen Psychiatry* (1996) 53:585–94. doi:10.1001/archpsyc.1996.01830070031007
40. Moorhead TWJ, Gountouna V-E, Job DE, McIntosh AM, Romaniuk L, Lymer GKS, et al. Prospective multi-centre voxel based morphometry study employing scanner specific segmentations: procedure development using CaliBrain structural MRI data. *BMC Med Imaging* (2009) 9:8. doi:10.1186/1471-2342-9-8
41. Brunton S, Gunasinghe C, Jones N, Kempton MJ, Westman E, Simmons A. A voxel-based morphometry comparison of the 3.0T ADNI-1 and ADNI-2 volumetric MRI protocols. *Int J Geriatr Psychiatry* (2015) 30:531–8. doi:10.1002/gps.4179
42. Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* (2002) 15:870–8. doi:10.1006/nimg.2001.1037
43. Rummel C, Slavova N, Seiler A, Abela E, Hauf M, Burren Y, et al. Surface based morphometry in individual patients with temporal lobe epilepsy. *Sci Rep* (2017) 7:10883. doi:10.1038/s41598-017-10707-1
44. Bishop CA, Newbould RD, Lee JS, Honeyfield L, Quest R, Colasanti A, et al. Analysis of ageing-associated grey matter volume in patients with multiple sclerosis shows excess atrophy in subcortical regions. *Neuroimage Clin* (2017) 13:9–15. doi:10.1016/j.nicl.2016.11.005
45. Steenwijk MD, Geurts JJ, Daams M, Tijms BM, Wink AM, Balk LJ, et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain* (2016) 139:115–26. doi:10.1093/brain/awv337
46. Jones SE, Buchbinder BR, Aharon I. Three-dimensional mapping of cortical thickness using Laplace's equation. *Hum Brain Mapp* (2000) 11:12–32. doi:10.1002/1097-0193(200009)11:1<12::AID-HBM20>3.0.CO;2-K
47. Das SR, Avants BB, Grossman M, Gee JC. Registration based cortical thickness measurement. *Neuroimage* (2009) 45:867–79. doi:10.1016/j.neuroimage.2008.12.016
48. Clarkson MJ, Cardoso MJ, Ridgway GR, Modat M, Leung KK, Rohrer JD, et al. A comparison of voxel and surface based cortical thickness estimation methods. *Neuroimage* (2011) 57:856–65. doi:10.1016/j.neuroimage.2011.05.053

**Conflict of Interest Statement:** AC received research support from Genzyme and Novartis well as personal compensation for activities with Almirall, Bayer, Biogen, Genzyme, Merck, Novartis, Sanofi, and Teva. AS received personal compensation for activities with Novartis, Sanofi, and Almirall Hermal GmbH. None of the other authors has any conflict of interest to disclose.

Copyright © 2018 Rummel, Aschwanden, McKinley, Wagner, Salmen, Chan and Wiest. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.