


# Benefits of a mosaic approach for assessing cortical atrophy in individual multiple sclerosis patients

Marlene Tahedl<sup>1</sup>  | Tun Wiltgen<sup>2</sup> | Cui Ci Voon<sup>2</sup> | Achim Berthele<sup>2</sup> |  
Jan S. Kirschke<sup>1</sup> | Bernhard Hemmer<sup>2</sup> | Mark Mühlau<sup>2</sup> | Claus Zimmer<sup>1</sup> |  
Benedikt Wiestler<sup>1</sup>

<sup>1</sup>Department of Neuroradiology, School of Medicine, Technical University of Munich, Munich, Germany

<sup>2</sup>Department of Neurology, School of Medicine, Technical University of Munich, Munich, Germany

## Correspondence

Marlene Tahedl, Department of Neuroradiology, School of Medicine, Technical University of Munich, Ismaninger Str. 22, Munich D-81675, Germany. Email: [marlene.tahedl@tum.de](mailto:marlene.tahedl@tum.de)

## Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: TA 1902/1-1

## Abstract

**Objective:** Cortical gray matter (GM) atrophy plays a central role in multiple sclerosis (MS) pathology. However, it is not commonly assessed in clinical routine partly because a number of methodological problems hamper the development of a robust biomarker to quantify GM atrophy. In previous work, we have demonstrated the clinical utility of the “mosaic approach” (MAP) to assess individual GM atrophy in the motor neuron disease spectrum and frontotemporal dementia. In this study, we investigated the clinical utility of MAP in MS, comparing this novel biomarker to existing methods for computing GM atrophy in single patients. We contrasted the strategies based on correlations with established biomarkers reflecting MS disease burden.

**Methods:** We analyzed T1-weighted MPRAGE magnetic resonance imaging data from 465 relapsing-remitting MS patients and 89 healthy controls. We inspected how variations of existing strategies to estimate individual GM atrophy (“standard approaches”) as well as variations of MAP (i.e., different parcellation schemes) impact downstream analysis results, both on a group and an individual level. We interpreted individual cortical disease burden as single metric reflecting the fraction of significantly atrophic data points with respect to the control group. In addition, we evaluated the correlations to lesion volume (LV) and Expanded Disability Status Scale (EDSS).

**Results:** We found that the MAP method yielded highest correlations with both LV and EDSS as compared to all other strategies. Although the parcellation resolution played a minor role in terms of absolute correlations with clinical variables, higher resolutions provided more clearly defined statistical brain maps which may facilitate clinical interpretability.

**Conclusion:** This study provides evidence that MAP yields high potential for a clinically relevant biomarker in MS, outperforming existing methods to compute cortical disease burden in single patients. Of note, MAP outputs brain maps illustrating individual cortical disease burden which can be directly interpreted in daily clinical routine.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Brain and Behavior* published by Wiley Periodicals LLC.

## KEYWORDS

biomarker, gray matter atrophy, individual MRI, MAP, mosaic-based approach, multiple comparisons, relapsing-remitting multiple sclerosis, smoothing

## 1 | INTRODUCTION

### 1.1 | Relevance of appreciating cortical disease burden in multiple sclerosis patients and the “MAP” approach

Multiple sclerosis (MS) is a debilitating disorder of the central nervous system (CNS) which remains incurable until today (Lassmann, 2018). The most prominent pathological substrate of MS pathology are lesions to the white matter (WM) of the CNS, which present as hyperintensities on T2-weighted FLAIR magnetic resonance images (MRI). Although WM lesion load does explain variance in terms of clinical disease burden—in particular cognitive functioning (Engl et al., 2020; Mollison et al., 2017) and its progression are negatively correlated with cognitive decline (Todea et al., 2020)—it has become clear that it does not capture the entirety of the clinical state (Barkhof, 2002; Popescu et al., 2013). Advances in neuroimaging methods allowed to identify another and at least partly independent anatomical substrate, namely, the shrinking and/or decay of CNS cells, or atrophy (Eshaghi et al., 2018; Filippi & Rocca, 2010; Vrenken & Geurts, 2007). Atrophy has been shown to be even more closely related to disability progression than WM lesion load (Bakshi et al., 2001; de Stefano et al., 2003; Sailer et al., 2003).

In our previous work, we have suggested to estimate individual gray matter (GM) cortical atrophy based on rating single subject's parceled cortical thickness (CT) data and evaluating the single “mosaics” with respect to healthy controls (HC) (Tahedl, 2020). In a series of studies, we have evaluated and demonstrated the utility of mosaic approach (MAP) for assessing clinically relevant cortical disease burden in motor neuron diseases (MND) and frontotemporal dementia (FTD) (see, Section 1.4). In this study, we aim to probe the clinical utility of MAP in MS, comparing this novel biomarker to existing methods for computing GM atrophy in single patients.

### 1.2 | Assessing individual deep gray matter and white matter disease burden

Atrophy of both deep GM nuclei as well as cortical atrophy reflect aspects of MS disability but might cover distinct features (Eijlers et al., 2019). However, both forms of atrophy do inform the clinical state, such that taking into account both markers—in addition to the well-established biomarker WM lesion load, as discussed above—should provide a more complete assessment of the individual patient's clinical state and might allow for more accurate prognoses. For effective clinical translation, such a biomarker needs to be readily acquirable and interpretable in a personalized manner.

For WM lesions and deep GM atrophy, the assessment of an individual estimate is relatively straightforward using advanced neuroimaging methods: In order to quantify WM lesion load (and thus allowing long-term monitoring), (half-)automated software-supported methods have been developed that can be applied to single subjects' T1w/T2-FLAIR data sets which output volume estimates of the individual lesion load with high accuracy (Mortazavi et al., 2012; Schmidt et al., 2012). Similarly, automated segmentation methods exist that provide reliable volume estimates for distinct deep GM nuclei (Fischl et al., 2002, 2004; Mendelsohn et al., 2023).

### 1.3 | Assessing individual cortical disease burden

Unlike WM lesion load and subcortical volumetry, GM atrophy is much harder both to quantify and to interpret on a single-patient level (Amiri et al., 2018). In terms of quantification, one problem is that cell damage or death induces a complex process of tissue restructuring which can hardly be reflected by a singular metric (Tsouki & Williams, 2021). Therefore, the neuroimaging community has suggested a wide number of metrics to quantify cortical GM changes, including the absolute CT (e.g., Fischl & Dale, 2000), its volume (e.g., Gaser et al., 2022) as well as “hybrids” such as GM concentration (most famously via the voxel-based morphometry [VBM], VBM, approach, e.g., Ashburner & Friston, 2000). All of these metrics have been demonstrated to mirror disease burden in MS on the group level; nevertheless, attempts to engineer metrics from them that can be applied to single subjects remain a challenge (Amiri et al., 2018).

### 1.4 | Challenges of existing methods for single-patient assessment of GM atrophy

The problem of constructing clinically relevant biomarkers for cortical atrophy is aggravated by the fact that neuroimaging pipelines regularly offer to adjust a wide range of parameters in both pre- and postprocessing of data which can impact downstream analysis results (Cash et al., 2015; Gunter et al., 2003; Popescu et al., 2016) and impede widespread adoption and generalization. The rationale behind many of those parameters is to increase statistical power of the analysis—it needs to be emphasized that in neuroimaging, the cortex is typically represented as hundreds of thousands of distinct data points. This requires rigorous control of the multiple comparisons problem, which can easily hamper statistical power if using common correction methods such as suggested by Bonferroni (Simes, 1986). One way to address this problem is to smooth the data before running statistics to reduce the variability between neighboring data points (Jo et al., 2007)—

however, notice that there is no clear consensus on the extent of “optimal” smoothing. Another strategy is to choose from less rigorous family-wise error rate (FWER) correction methods, tailored to neuroimaging data, such as threshold-free cluster enhancement (TFCE, Smith & Nichols, 2009) or employ alternative approaches, for example, the false discovery rate (FDR, Benjamini & Hochberg, 1995; Genovese et al., 2002).

## 1.5 | Employing a mosaic approach (MAP) to assess clinically relevant cortical disease burden

The above list of choices to identify cortical GM changes is far from being complete but indicates how one might easily get overwhelmed by the abundant number of possible combinations throughout data analysis. Moreover, there is no clear directive on how adjustment of the technical details and their combinations impacts downstream results, which in turn hampers interpretability. This might be one reason why the development of a readily accessible biomarker for interpreting single-patient GM atrophy is still lacking in MS.

To overcome this gap, we have developed the “MAP” (Tahedi, 2020), which assesses individual cortical disease burden for distinct brain regions (“mosaics”). The primary outcome variable of MAP is a single number reflecting the fraction of potentially atrophic cortex (referred to as “thin-patch fraction”). In a series of studies, we have evaluated and demonstrated the utility of MAP for assessing clinically relevant cortical disease burden in MND and FTD: More specifically, we could demonstrate that MAP reflects both the cross-sectional as well as the longitudinal progression of clinical disability in amyotrophic lateral sclerosis (ALS), as measured by the ALS Functional Rating Scale, revised (Tahedi et al., 2021). Moreover, we could show that MAP accurately reflects the histopathologic involvement of cortical disease burden: By contrasting (1) a purely upper motor neuron—that is, purely cortical—condition, namely, primary progressive lateral sclerosis (PLS), (2) a mixed upper-/lower motor neuron condition (ALS), and (3) a primary lower motor neuron condition (poliomyelitis survivors), we could show that this degree of cortical involvement is captured by MAP—both cross-sectionally and longitudinally (Tahedi et al., 2022). Furthermore, we have investigated the clinical utility of MAP outside of MND and applied it to the FTD spectrum, where we could also demonstrate its clinical utility for single patients (McKenna et al., 2022).

## 1.6 | MAP—a novel biomarker for assessing individual atrophy also in MS?

Our previous investigations into MND and FTD motivate us to interrogate the utility of MAP also in MS, given the need for an individual cortical biomarker in that condition as outlined above. In this study, we provide a thorough investigation of MAP utility in MS. Importantly, we focus our study on comparisons with existing strategies to compute individual cortical disease burden (which we refer to as “standard approaches”) and technical variations of those (including

smoothing and statistical correction methods, Figure 1). We also consider variations of MAP by contrasting different parcellation schemes and varying the resolution of the respective partitioning. We appreciate both effects on individual- and group-level statistical comparisons between the methods. Moreover, we validate the clinical utility of the approaches based on their associations with correlations to third metrics reflecting clinical disease burden in MS, namely, WM lesion load and Expanded Disability Status Scale (EDSS) scores, which both reflect at least partly independent and complementary information on the MS disease course/progression (Sormani et al., 2014). With this careful inspection, we aim to explore the benefits of MAP for assessing clinically relevant cortical disease burden for single MS patients, thereby extending MAP generalizability beyond MND and FTD. Indeed, MAP outputs brain maps that illustrate single-subject cortical involvement, which can be used for personalized assessment and monitoring in MS clinical routine. We end our investigation with a demonstration of such single-subject visualization.

## 2 | METHODS

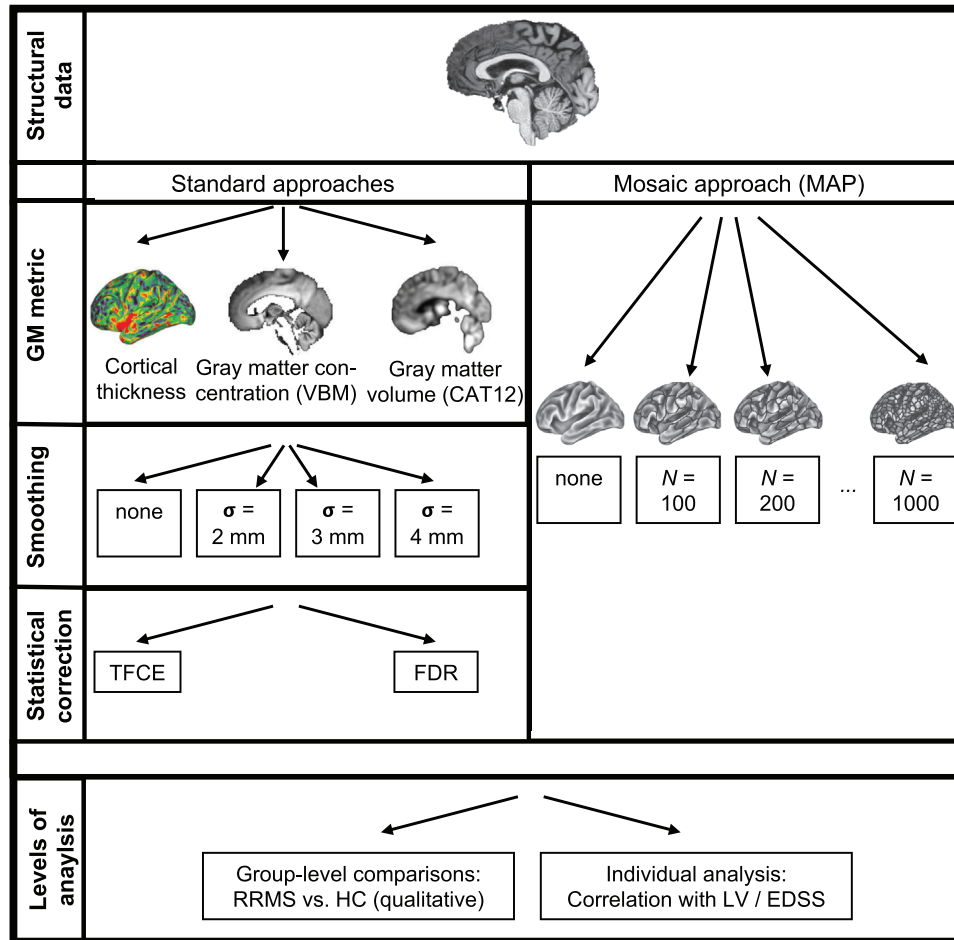
### 2.1 | Subjects

In total, 465 relapsing-remitting MS (RRMS) patients and 89 HC were included in this study (Table 1). Patients were diagnosed according to the 2017 revisions of the MacDonald criteria (Thompson et al., 2018). 3D T1w MPRAGE images were acquired at the Klinikum rechts der Isar of the Technical University of Munich on one of two 3.0 T whole body Philips MRI scanners (1) *Achieva dStream*, (2) *Ingenia* with identical scanning parameters: 267 sagittal slices, field of view = 240 × 252 mm<sup>2</sup>, spatial resolution = 1.00 mm<sup>3</sup>, repetition time = 9 ms, echo time = 4 ms, flip angle = 8°, and acquisition time = 2 min 25 s. Moreover, demographic data, including age and gender were recorded. For patients, additionally disease duration, EDSS scores and the dominant hand were assessed at a time point close to the MR session (±1 week) by a certified neurologist. All subjects had given informed consent to the use of their data for research purposes. The study procedures followed safety guidelines for MRI research at the Technical University of Munich, which are in line with the Declaration of Helsinki and were approved by the local ethics committee. Demographic variables including age means and gender distributions between RRMS and HC were performed using Welch’s two-sample *t*-tests and Chi-squared test with Yate’s continuity correction.

### 2.2 | Neuroimaging data analysis

#### 2.2.1 | Preprocessing

The idea of this study was to evaluate and compare three different commonly used strategies for assessing GM variations, as well as to document impacts of smoothing and methods to correct for the multiple comparisons problem (Figure 1). As such, we preprocessed



**FIGURE 1** Study objective. The current work contrasted different strategies for assessing gray matter (GM) variations in relapsing-remitting multiple sclerosis (RRMS) patients. In specific, we compared different metrics for calculating GM differences (cortical thickness [CT], GM concentration via VBM, GM volume via CAT12), smoothing (no smoothing,  $\sigma = 2/3/4$  mm), and correction methods for multiple comparisons (threshold-free cluster enhancement [TFCE]- vs. false discovery rate [FDR]-correction). In addition, we considered the mosaic approach (MAP), a method to assess CT differences based on parcellation the cortex into smaller subregions and, for individual patients, calculating whether or not that parcel was significantly thin as compared to a healthy control (HC) control group. We evaluated both group-level comparisons (by inspecting statistical output maps between the combinations) as well as individual-level comparisons (by investigating which combination correlated best with lesion volume [LV] on the one hand and expanded disability status score [EDSS] on the other hand). CAT12, computational anatomy toolbox, version 12 (used for calculating GM volume); VBM, voxel-based morphometry (used for estimating GM concentration).

structural data for all subjects using three different analysis pipelines to calculate (a) CT, (b) VBM, which reflect aspects related to GM concentration, and (c) GM volume (using the computational anatomy toolbox [CAT12]). To calculate CT, we started with a full image segmentation and surface reconstruction using FreeSurfer's *recon-all* pipeline (Dale et al., 1999; Fischl, 2012; Fischl et al., 1999). These data were then converted to the CIFTI format to improve visualization and downstream data handling (Dickie et al., 2019). The main output from this pipeline for our study was a file which estimates CT at each of 32k vertices per hemisphere for each subject. VBM was calculated within FMRIB's Software Library (FSL) (Jenkinson et al., 2012), after running its structural processing algorithm *fs\_l\_anat*. We calculated voxelwise VBM maps for each patient within FSL (using the software's standard recommendations, Good et al., 2002). Finally, GM volume was assessed using the CAT12 toolbox from the SPM12 software package, also fol-

lowing the manual's standard guidelines (Schmidt et al., 2012). Both VBM and CAT12 data were then warped to the MNI152 2 mm standard space to enhance group-level comparisons. Finally, WM lesion segmentation was performed using the lesion growth algorithm of LST toolbox, which outputs an estimate of a given subject's total lesion volume (LV), among others.

### 2.3 | Smoothing

One goal of this study was to document effects of different smoothing factors on the preprocessed data on statistical comparisons. In brief, the rationale behind smoothing is to account for inaccuracies in spatial registration as well as increase the signal-to-noise ratio (SNR), e.g., Hopfinger et al., 2000, who published a comprehensive analysis of

**TABLE 1** Demographic details of the study populations.

	RRMS patients	HC	Group differences (W <sup>+</sup> /C <sup>2++</sup> )
Total number of subjects	465	89	n.a.
Age [y, mean ± SD]	40.14 ± 9.94	37.36 ± 15.06	W: $t(103.16) = 1.67$ , $p = .10$
Sex, F/M	305/160	59/30	C <sup>2</sup> : $\chi^2(1, N = 554) = 3.27e - 5$ , $p = .995$
Dominant hand, R/L	408/57	n.a.	n.a.
Average whole brain cortical thickness [mm, mean ± SD]	2.35 ± 0.09	2.41 ± 0.11	W: $t(113.31) = -4.60$ , $p < .001^*$
Lesion volume [mL, mean ± SD]	5.39 ± 8.41	n.a.	n.a.
EDSS score [median (IQR)]	1.50 (2.00)	n.a.	n.a.
Disease duration [y, mean ± SD]	8.65 ± 5.16	n.a.	n.a.

Note: <sup>+</sup> = Welch two-sample *t*-tests were performed to test differences of age and years of education between all patients versus HC, <sup>++</sup> = Chi-square tests were performed to test differences of sex and handedness frequencies between all patients versus HC.

Abbreviations: C<sup>2</sup>, Chi-square test; EDSS, expanded disability status score; F, female; HC, healthy control; L, left-handed; M, male; ml, milliliter; MRI, magnetic resonance imaging; MS, multiple sclerosis; N, sample size; n.a., not applicable; R, right-handed; RRMS, relapsing-remitting MS; SD, standard deviation; W, Welch two-sample *t*-test; y, years.

\*Significant at an alpha-level of  $p \leq .05$ .

smoothing effects on fMRI data). In total, we contrasted four different smoothing factors, namely, (a) no smoothing ( $\sigma = 0$  mm), (b)  $\sigma = 2$  mm, (c)  $\sigma = 3$  mm, and (d)  $\sigma = 4$  mm. Notice that commonly, smoothing factors are provided as full-width at half maximum (FWHM) Gaussian kernels, which can be directly calculated from sigma by multiplying with a factor of approximately 2.3548, such that, for example,  $\sigma = 2$  mm translates into a Gaussian kernel FWHM = 4.7096 mm. We applied those smoothing factors to the outputs of all three structural preprocessing pipelines, whereas we used tools from Workbench (Glasser et al., 2013; Marcus et al., 2011) for the surface-based CT data and tools from FSL (Jenkinson et al., 2012) for smoothing volumetric-based data, that is, VBM and CAT.

## 2.4 | Cortical parcellation

As specified above, one rationale behind smoothing is to boost SNR and therefore statistical power. Alternatively, one can increase power by reducing the number of statistical comparisons, for example, by averaging across neighboring data points (Tahedi, 2020). This “parcellation method” has been suggested previously for CT data and has been demonstrated to yield clinically relevant information for various neurological conditions such as ALS (Tahedi et al., 2021) and FTD (McKenna et al., 2022). However, these studies did not interrogate effects of different CT parcellation schemes but have used a cortical parcellation of  $N = 1000$  “patches” (in total for both hemisphere), using predefined and roughly equally sized atlas regions (Schaefer et al., 2018). Here, we considered 10 different parcellation schemes (from  $N = 100$  to 1000 patches in steps of  $N = 100$ ), as well as no parcellation (i.e., each vertex is its own “patch,” such that  $N = 59,234$  patches). The parcellation

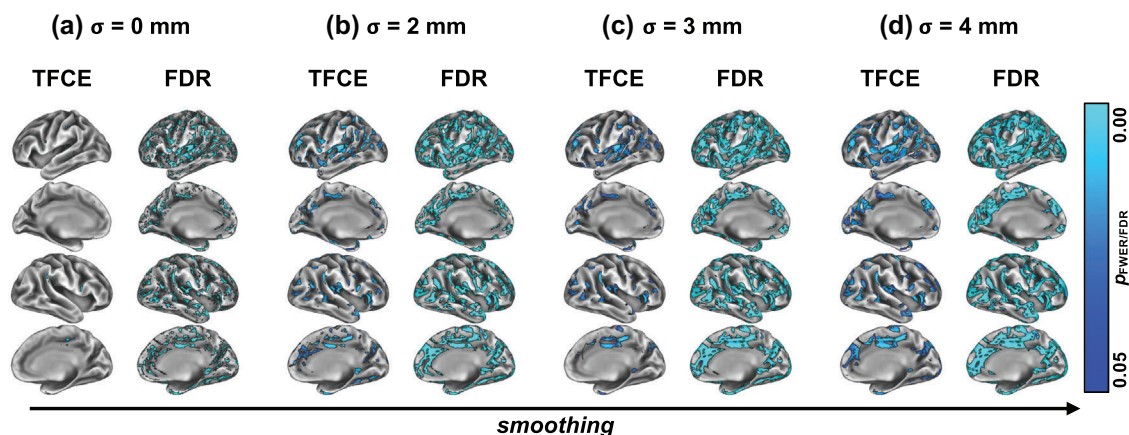
schemes were taken from work by Schaefer et al. (2018), analogous to the original studies suggesting the parcellation method.

## 2.5 | Group-level statistics

### 2.5.1 | Study design: “Standard approaches”

We were interested in effects of different “standard approaches” to assess GM variations both on group-level statistical comparisons as well as individual-versus-group statistical comparisons. All statistics (but for the CT parcellation, see below) were run within FSL using its *randomize* tool (notice that CT data were converted temporarily to NIFTI data using tools from Workbench to ensure comparability with FSL algorithms). For all three standard approaches of assessing GM alterations, we limited our contrasts to a GM cortical mask (see Figure 2A). We used general linear models (GLM), corrected for age and gender, for both group-level and individual-versus-group statistics. For the group-level statistics, we set up the design matrix as a two-group difference comparison, whereas for the individual-versus-group comparisons we set up the design matrix as “Singleton-versus-Group” or “Prediction-Interval-Test” (following FSL’s recommendations/terminology on setting up GLMs), which is basically the same test as specified for the group contrast when one group has exactly one subject in it, that is, in our case the respective RRMS patient (which was tested against  $N = 89$  HC and run for each patient separately). Both comparisons were adjusted for the covariates age and gender. As we were merely interested in GM variations indicating atrophy, we ran one-sided testing and considered only the contrast  $MS < HC$ .





**FIGURE 2** Group-level comparison of cortical thickness (CT). Contrasting CT between relapsing-remitting multiple sclerosis (RRMS) and healthy control (HC) for different neuroimaging parameters, we found that family-wise error (FWE)-correction yields “cleaner” output maps as compared to false discovery rate (FDR)-corrected statistical maps (compare left vs. right columns on A–D). In addition, no smoothing (a) almost suppressed any significant results for FWE-correction, smoothing kernels of  $\sigma = 2$  mm (b), 3 mm (c), and 4 mm (d) produced clearly defined maps of gray matter (GM) variations, centered around the left, and right insular/temporal cortices. TFCE, threshold-free cluster enhancement.

## 2.5.2 | Correction for multiple comparisons: “Standard approaches”

We contrasted two common ways to account for multiple testing for standard approaches, namely, (a) TFCE and (b) FDR. Although TFCE can be directly applied with FSL’s *randomize* (notice that we used the algorithm’s default settings, that is, 5000 permutations, among others), FDR-correction can be run on the *p*-maps uncorrected for multiple comparisons, which can be also output from *randomize*. Also, for latter correction, we used FSL’s default settings (i.e., voxelwise-thresholding at  $p \leq .05$ , 5000 permutations).

## 2.5.3 | Study design: mosaic approach (MAP)

The study design for the MAP, that is, parcellating CT data and evaluating it with respect to HC, deviated from the standard approaches given the rationale behind this strategy (which was to assess whether a single subject’s patch is “significantly thin” as compared to a reference group). As suggested in the original MAP paper (Tahedl, 2020), we started by averaging all vertices for each patch (of all 10 parcellations) and subjects. The HC group served as the reference group. Then, each subject’s patch (both for patients and HC) was compared against that reference group using nonparametric statistical permutation testing with in-house software written in MATLAB, version R2022b (The MathWorks), correcting for FWE (Nichols & Holmes, 2002). In brief, we started by computing *z*-scores for each patch and patient, by rating each of a given patient’s patch versus those of the HC group. These *z*-scores were then converted to *p*-values using a permutation procedure to correct for the FWER: In brief, we shuffled the data (combining the patient and their controls) under the null hypothesis of no difference between HC and patient and calculated the respective *z*-scores on each iteration (for each patch separately to account for the physiolog-

ical differences of thickness across the cortex, Tahedl, 2020). For each patient, we applied an exhaustive permutation procedure, combining all 89 HC and the respective patient. This resulted in a personalized, nonparametric null distribution for each patient and patch. The *p*-value was then defined as ratio between all *z*-scores in the null distribution smaller than the observed patient’s *z*-score and the number of iterations. We set the  $p_{\text{FWER}}$ -value to  $p_{\text{FWER}} \leq .05$ .

## 2.6 | Assessing clinical relevance for individual data

One major focus of this study was to evaluate which of the analyzed strategies for assessing GM variations yielded clinically relevant information on an individual patient level. As such, we set up a correlation analyses which analyzed the association between the respective analysis strategy (i.e., each combination of GM metrics, smoothing factor and correction method, as well as MAP) and two clinical variables that cover both structural disease burden (by incorporating LV as DV), as well as functional deficits (by considering EDSS scores as DV). We set up separate correlation analyses for the GM metrics/clinical variables. Notice that to quantify GM changes for a single subject, we used a scalar for each patient as metric of interest which was simply the ratio of all voxels/vertices (smoothed data) or patches (MAP) which were found to be significantly “thin”—suggestive of atrophy—by the above-specified methods (notice that the reference for that ratio calculation was the cortical mask shown in Figure 2a).

To assess differences between the alternative methods employed, we ran analyses of variance (ANOVA) in each step of the analysis stream as shown in Figure 1, that is, we tested three specific hypotheses with regard to correlation with the dependent clinical variables LV/EDSS: (1) GM changes are not equally correlated to LV and EDSS (testing for differences of correlations with the dependent

variables, regardless of the strategy for computing GM changes). (2) There is a difference of the “approach” (testing for differences of [any] standard approach vs. [any] MAP parcellation scheme). (3) Smoothing/parcellation in general makes a difference (testing no smoothing vs. any smoothing for the standard approaches and no parcellation vs. any parcellation for MAP).

Where applicable, we followed up significant ANOVA results in post hoc pairwise comparison using Tukey’s Honest Significant Difference (HSD) tests with adjusted  $p$ -values, correcting for FWER. These correlation analyses, ANOVAs, and Tukey HSD tests were all run within *RStudio* (using R version 4.2.2, 2022, RStudio Team, 2022), which was also used for figure conception.

## 2.7 | Data availability statement

All presented group-level outputs, statistical maps, variable distributions, post hoc statistics, and additional information on data processing pipelines can be requested from the corresponding author. However, individual patient clinical and neuroimaging data cannot be made available.

## 3 | RESULTS

### 3.1 | Demographics

Adequate age-matching of the RRMS and HC group was suggested by a nonsignificant comparison of age means between the two groups using the Welch two-sample  $t$ -test ( $t(103.16) = 1.67, p = .10$ , see Table 1). Similarly, the distribution of gender among the study groups did not change significantly in a Chi-square test ( $\chi^2(1, N = 554) = 3.27e^{-5}, p = .995$ ). The mean disease duration in the RRMS group was 8.65 years ( $SD \pm 5.16$ ). RRMS patients had thinner cortices as opposed to HC ( $t(113.31) = -4.60, p < .001$ ). LV was on average 5.39 mL ( $SD \pm 8.41$  mL) in the RRMS group and the median EDSS score was 1.50 (IQR = 2.00), suggesting a relatively preserved patient group from a clinical perspective.

### 3.2 | Qualitative group comparisons suggest high sensitivity of cortical thickness to detect gray matter changes

Comparing the group contrasts qualitatively between the RRMS and HC groups with respect to the metric used for assessing GM atrophy (i.e., CT: Figure 2, VBM: Figure 3, CAT12: Figure 4), we noted that CT (Figure 2) resulted in most data points of significant changes in the RRMS group throughout the cortex, for all levels of smoothing. VBM suggests GM changes in a very sharply defined region around the right insula (Figure 3), whereas CAT12 failed to yield any significantly different data points for the default alpha threshold (except for a few single

voxels at the right insular cortex for no smoothing, Figure 4). Only at a more liberal alpha threshold ( $p \leq .20$ ) did we find “significant” voxels between the study groups, which roughly matched with the location found in the VBM contrast.

### 3.3 | Focused gray matter changes of the right temporal/insular cortices are found using different methods

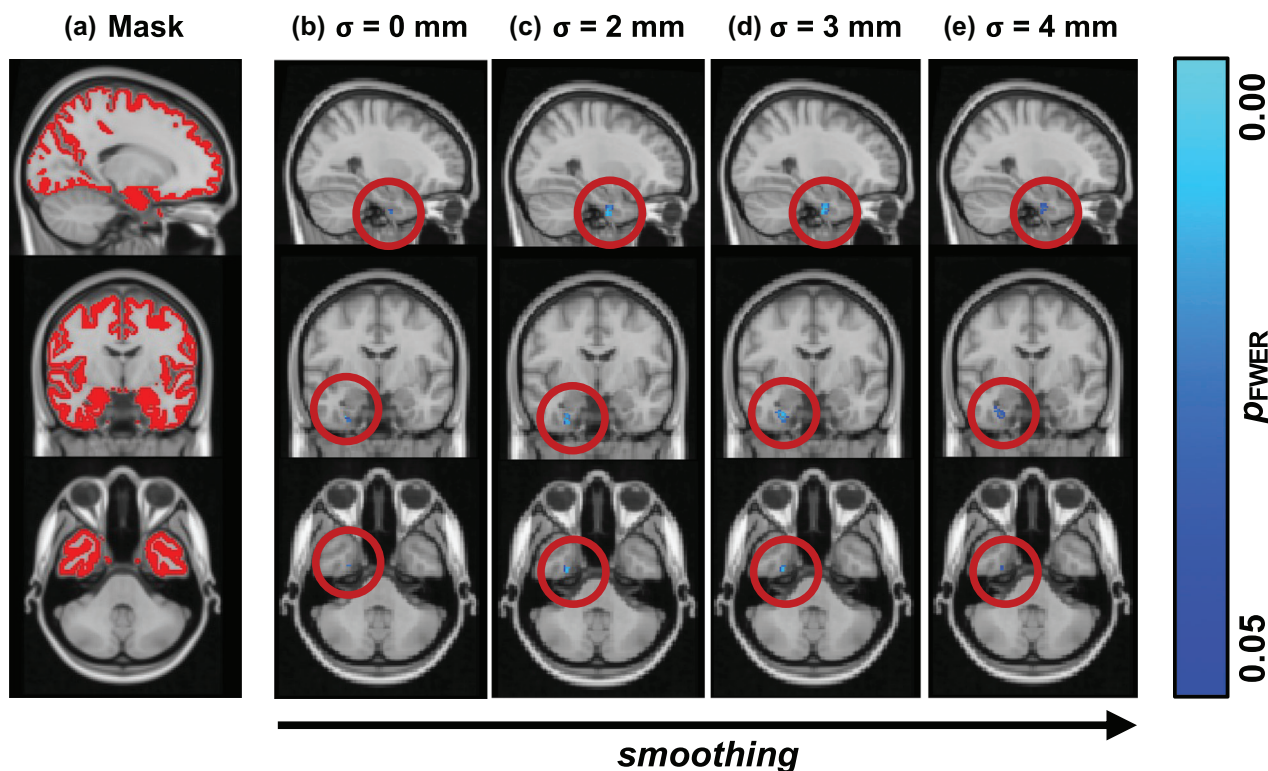
As noted above, the VBM contrast (Figure 4) suggested a very sharply defined affected GM region in the RRMS group around in right insular cortex. Similarly, the liberally thresholded CAT12 contrast yielded a region close to the one described in VBM with a slightly more lateral location, making borders with the temporal lobe. Also, especially the TFCE-corrected results of the CT comparison yield focused GM changes around the bilateral temporal/insular cortices; however, the affected areas found with CT contrasts covered wide parts of the cortex, including bilateral parietal, frontal, and visual areas.

### 3.4 | Smoothing enhances cleanliness but might have an optimum

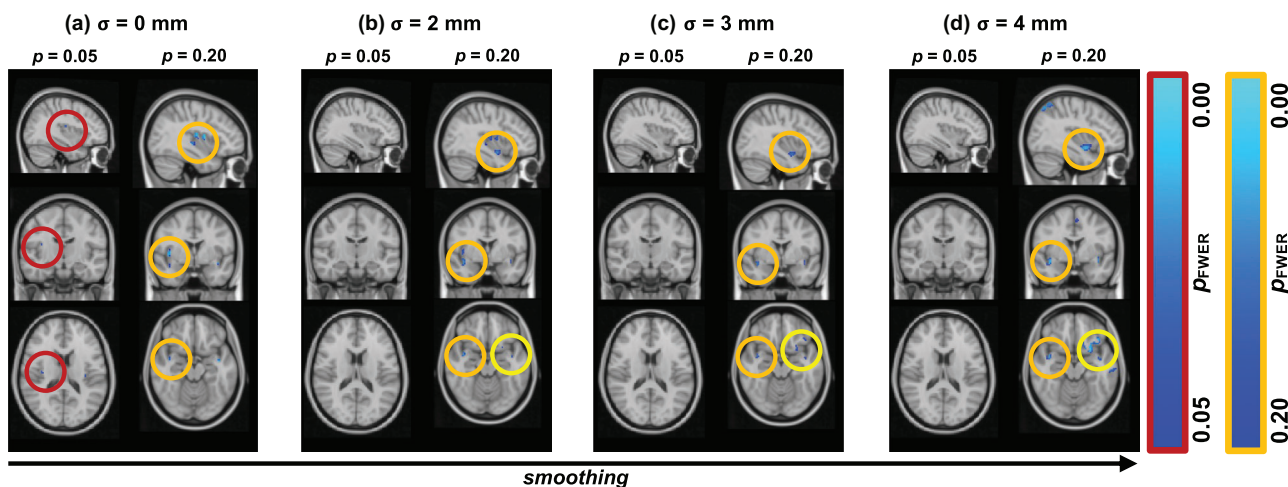
In terms of smoothing, the qualitative comparisons of the different group contrasts (cf. columns in Figures 2–4, respectively) suggested mainly two things: (1) Smoothing, in general, seems to slightly enhance statistical power in that it results in a greater number of statistically different data points, regardless of the metric and/or correction method applied. (2) Statistical power and extent of smoothing seem not to be linearly correlated, that is, ever-more smoothing does not guarantee higher statistical power. This is evident, for example, in the VBM contrast comparing  $\sigma = 3$  mm, Figure 3c, versus  $\sigma = 4$  mm, Figure 3d. When using  $\sigma = 4$  mm, the significant result constituted fewer data points, and those data points yielded higher  $p$ -values than for  $\sigma = 3$  mm. This suggests an optimum of smoothing with loss of statistical power for both increases and decreases of smoothing extent from that optimum.

### 3.5 | TFCE-correction outputs relatively more and cleaner results as compared to FDR

Comparing correction methods for multiple comparisons on assessing GM differences between the groups, we noted that TFCE seems superior in terms of statistical power. Two main observations support this claim: (1) FDR-corrected  $p$ -values for VBM and CAT12 data did not yield any significant data points at all (and are therefore not shown in Figures 3 and 4) and (2) when FDR-correction did show significant differences as with CT data (Figure 2, right columns), the results were much more spread across the cortex and not as clearly defined as TFCE-corrected results for all levels of smoothing.

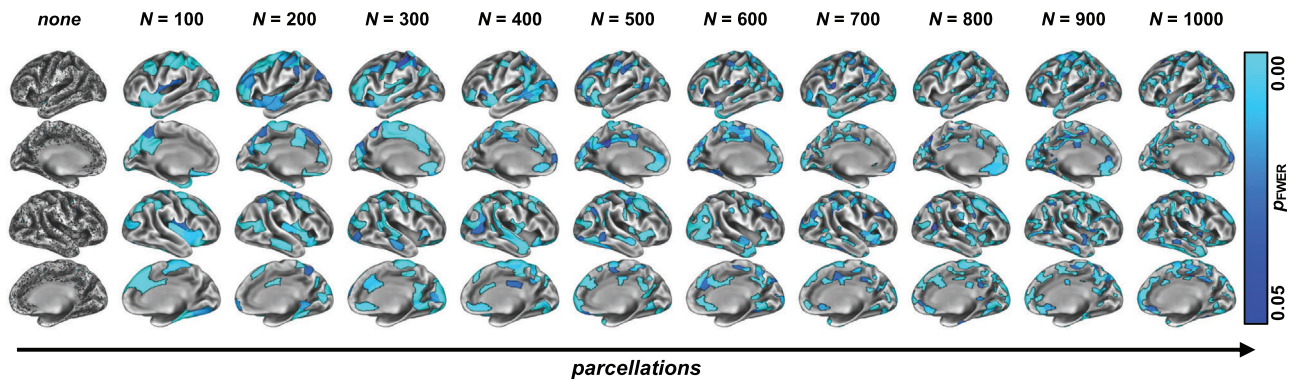


**FIGURE 3** Group-level comparison of voxel-based morphometry (VBM). Contrasting VBM between relapsing-remitting multiple sclerosis (RRMS) and healthy control (HC) for different neuroimaging parameters, we found that false discovery rate (FDR)-correction (not shown) suppressed any significant results for all levels of smoothing. Therefore, we only show family-wise error (FWE)-corrected maps here. Part (a) shows the cortical mask to which we limited our statistical comparisons. All levels of smoothing suggested that a region in the deep right temporal lobe around the parahippocampal gyrus was affected in RRMS as compared to HC. However, although no smoothing (b) and extreme smoothing (e) showed relatively little significant voxels, moderate smoothing (c and d) yielded more clearly defined problematic regions in RRMS. The statistical maps are shown on the MNI152 2 mm standard template at voxel location 36-60-17. GM, gray matter; TFCE, threshold-free cluster enhancement.



**FIGURE 4** Group-level comparison of gray matter (GM) content. Contrasting CAT12 between relapsing-remitting multiple sclerosis (RRMS) and healthy control (HC) for different neuroimaging parameters, we found that false discovery rate (FDR)-correction (not shown) suppressed any significant results for all levels of smoothing. Therefore, we only show family-wise error (FWE)-corrected maps here. However, also FWE-correction suppressed almost any significant results at the alpha-level  $p \leq .05$  (left columns in a-d). The only exceptions were a few significant voxels for no smoothing (a) at the right insular cortex (red circle). As we chose a more liberal threshold ( $p \leq .20$ , right columns in a-d), we found this same region to be affected for all levels of smoothing, supplemented by more regions in the right temporal/insular cortex (orange circles). However, notice that more smoothing (d) also introduced noisy regions (cf. yellow circles in the axial slices of b-c). The statistical maps are shown on the MNI152 2 mm standard template at voxel location 27-56-44 for  $p_{FWE} \leq .05$  (left columns) and at 25-63-28 for  $p_{FWE} \leq .20$  (right columns). CAT12, computational anatomy toolbox, version 12 (used for calculating GM content); TFCE, threshold-free cluster enhancement.





**FIGURE 5** Group-level comparison of the mosaic approach (MAP). We investigated MAP, a strategy to assess gray matter (GM) variation based on identifying, for single relapsing-remitting multiple sclerosis (RRMS) patients, which cortical parcels (averaged across all data points constituting that parcel) of a predefined parcellation scheme varied significantly from a healthy control (HC) control group. We ran nonparametric group-level statistics on these individual statistical maps for 10 different parcellation schemes as well as no parcellation (leftmost column). Although no parcellation produced a rather noisy map, the parceled data suggested GM damage in widespread cortical regions, focused around the bilateral temporal/insular cortices as well as the Somatomotor cortices. Moreover, higher resolution parcellation (i.e., moving rightwards in the figure) produced more clearly defined output maps.

### 3.6 | MAP yields qualitatively similar results as compared to standard approaches

As we inspected the group-level comparisons of MAP (Figure 5), we made three main observations: (1) As the resolution of the MAP parcellation increased (i.e., the more parcels the cortex was divided into), the results became finer (e.g., compare  $N = 100$  in Figure 5 with  $N = 1000$ ). However, no parcellation (i.e., treating each vertex as its own parcel, which is analogous to  $N = 59,234$  parcels or “no smoothing”) output a rather random map, which is clinically hard to interpret. Finally, (2) the topography of the affected region stays similar as the parcellation gets finer and is similar to the “standard approach” non-parceled CT group contrast (Figure 2), with affected regions throughout the cortex including temporal/insular cortices.

### 3.7 | MAP yields highest correlations with clinical variables

One main rationale of this study was to evaluate MAP versus standard approaches for their clinical relevance. We defined “clinical relevance” twofold, namely, (1) LV, to cover structural damage and (2) EDSS, to cover functional deficits. As we compared all combinations of GM metrics/smoothing/correction methods (Table 2), we made three main observations according to our hypotheses (Figure 6, Table 2): (1) In general, GM variation was stronger associated with LV as compared to EDSS score ( $F(3,66) = 41.32, p < .001$ ); compare general level of correlations values in Figure 6a vs. b. (2) MAP yielded higher correlations to both LV and EDSS as any of the standard approach combinations (ANOVA:  $F(3,66) = 24.64, p < .001$ ; post hoc Tukey HSD: MAP vs. CT:  $p_{\text{adj}} = .005$ , MAP vs. VBM:  $p_{\text{adj}} < .001$ , and MAP vs. CAT12:  $p_{\text{adj}} < .001$ ). (3) For the “standard approaches,” there was no statistical difference for smoothing ( $F(1,46) = .388, p = .54$ ). Similarly, for MAP, the ANOVA

suggested no significant difference for the specific parcellation scheme used ( $F(10,11) = .03, p = 1.00$ ).

## 4 | DISCUSSION

In the present study, we sought to compare different neuroimaging strategies for assessing cortical GM variations in RRMS patients as compared to HC. We centered our analysis around a previously suggested method to appreciate individual cortical disease burden in MND and FTD, namely, the MAP, which is based on estimating individual cortical burden with respect to referencing CT data with respect to HC. We evaluated the utility of MAP in MS by comparing its performance against existing approaches to compute individual cortical disease burden (“standard approaches”): Specifically, we investigated three aspects of calculating GM variability, namely, (1) the neuroimaging metric per se (we contrasted three commonly used GM metrics in the neuroimaging community, CT, and GM concentration (via VBM) and GM volume), (2) the smoothing factor (no smoothing and  $\sigma = 2/3/4$  mm), and (3) the correction method for multiple comparisons (FWER vs. FDR corrections). We ran group comparisons between RRMS and HC groups and qualitatively compared the resulting statistical maps. In addition to these “standard approaches,” we also considered MAP, a recently introduced method for assessing GM variation as CT differences on an individual patient’s level (Tahedi, 2020; Tahedi et al., 2021), which is based on subdividing the cortex into several subregions, then averaging across all data points constituting that parcel and hence checking for an individual patient whether or not that parcel’s observed CT significantly varies from a HC population. In total, we contrasted 10 such parcellation schemes (ranging from  $N = 100$  to 1000 parcels in steps of 100, as well as no parcellation). We computed group-level statistics on the individual patient’s parcellations using nonparametric statistical testing and inspected qualitative

**TABLE 2** Correlation values of individual study.

Outcome variable	Gray matter metric	Correction method	Smoothing (sigma)/ parcellation number	Correlation (Pearson's <i>r</i> )
LV	CT	FWE	0	.525
LV	CT	FWE	2	.513
LV	CT	FWE	3	.471
LV	CT	FWE	4	.469
LV	CT	FDR	0	.502
LV	CT	FDR	2	.49
LV	CT	FDR	3	.472
LV	CT	FDR	4	.456
LV	VBM	FWE	0	.265
LV	VBM	FWE	2	.485
LV	VBM	FWE	3	.457
LV	VBM	FWE	4	.405
LV	VBM	FDR	0	.209
LV	VBM	FDR	2	.453
LV	VBM	FDR	3	.401
LV	VBM	FDR	4	.364
LV	CAT12	FWE	0	.235
LV	CAT12	FWE	2	.224
LV	CAT12	FWE	3	.207
LV	CAT12	FWE	4	.205
LV	CAT12	FDR	0	.188
LV	CAT12	FDR	2	.142
LV	CAT12	FDR	3	.134
LV	CAT12	FDR	4	.137
LV	<b>MAP</b>	<b>Perm</b>	<b>0</b>	<b>.583</b>
LV	MAP	Perm	100	.505
LV	MAP	Perm	200	.529
LV	MAP	Perm	300	.53
LV	MAP	Perm	400	.537
LV	MAP	Perm	500	.54
LV	MAP	Perm	600	.547
LV	MAP	Perm	700	.552
LV	MAP	Perm	800	.549
LV	MAP	Perm	900	.558
LV	MAP	Perm	1000	.55
EDSS	MAP	Perm	0	.369
EDSS	MAP	Perm	100	.327
EDSS	MAP	Perm	200	.346
EDSS	MAP	Perm	300	.354
EDSS	MAP	Perm	400	.349
EDSS	MAP	Perm	500	.359
EDSS	MAP	Perm	600	.355
EDSS	MAP	Perm	700	.367

(Continues)

TABLE 2 (Continued)

Outcome variable	Gray matter metric	Correction method	Smoothing (sigma)/ parcellation number	Correlation (Pearson's <i>r</i> )
EDSS	MAP	Perm	800	.368
EDSS	<b>MAP</b>	<b>Perm</b>	<b>900</b>	<b>.374</b>
EDSS	<b>MAP</b>	<b>Perm</b>	<b>1000</b>	<b>.374</b>
EDSS	CT	FWE	0	.126
EDSS	CT	FWE	2	.137
EDSS	CT	FWE	3	.137
EDSS	CT	FWE	4	.132
EDSS	CT	FDR	0	.129
EDSS	CT	FDR	2	.129
EDSS	CT	FDR	3	.117
EDSS	CT	FDR	4	.129
EDSS	VBM	FWE	0	.158
EDSS	VBM	FWE	2	.271
EDSS	VBM	FWE	3	.245
EDSS	VBM	FWE	4	.234
EDSS	VBM	FDR	0	.156
EDSS	VBM	FDR	2	.235
EDSS	VBM	FDR	3	.196
EDSS	VBM	FDR	4	.212
EDSS	CAT12	FWE	0	.03
EDSS	CAT12	FWE	2	.028
EDSS	CAT12	FWE	3	.019
EDSS	CAT12	FWE	4	.028
EDSS	CAT12	FDR	0	-.023
EDSS	CAT12	FDR	2	.003
EDSS	CAT12	FDR	3	.001
EDSS	CAT12	FDR	4	.006

Note: Highest correlation values for EDSS and LV are marked in bold.

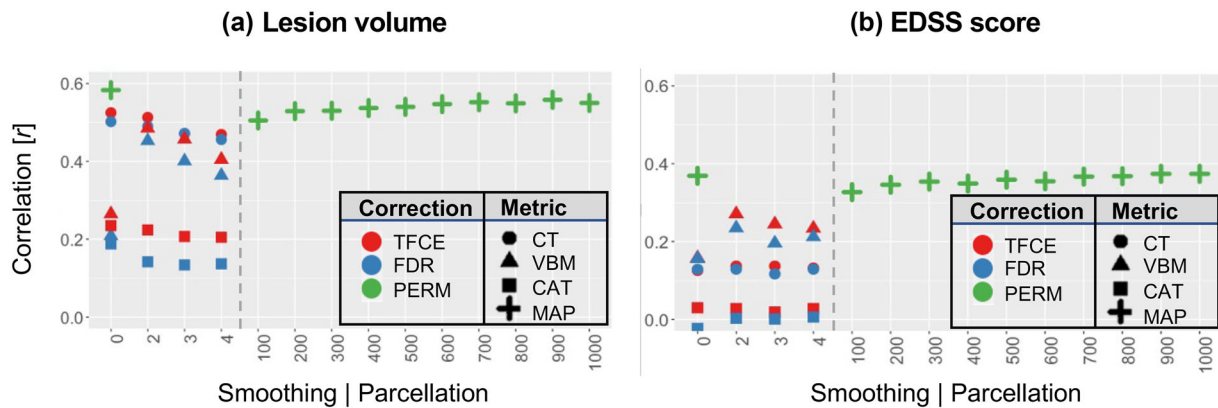
Abbreviations: CAT12, computational anatomy toolbox, version 12; CT, cortical thickness; EDSS, expanded disability status score; FDR, false discovery rate; FWE, family-wise error; LV, lesion volume; MAP, mosaic approach; Parc, parcellation; Perm, permutation; VBM, voxel-based morphometry.

differences between MAP and the “standard approaches” and effects of more high-resolution parcellation schemes. Finally, we investigated differences of the analyzed strategies (including all combinations of standard approaches as well as MAP) on correlations to clinically relevant outcome variables on for individual patient's, namely, LV and EDSS. For the (qualitative) group comparisons, we showed that CT produced most sensitive statistical maps, showing widespread GM changes in the RRMS group. VBM and CAT12 only indicated very focused changes around the temporal/insular cortices, which however was only detectable with a more liberal alpha threshold for some combinations. For the individual analysis, it turned out that the MAP method yielded higher correlations than any combination of the standard approaches for both LV and EDSS. However, LV was higher correlated to GM changes as compared to EDSS scores. Although an ANOVA suggested no differences for the specific parcellation scheme

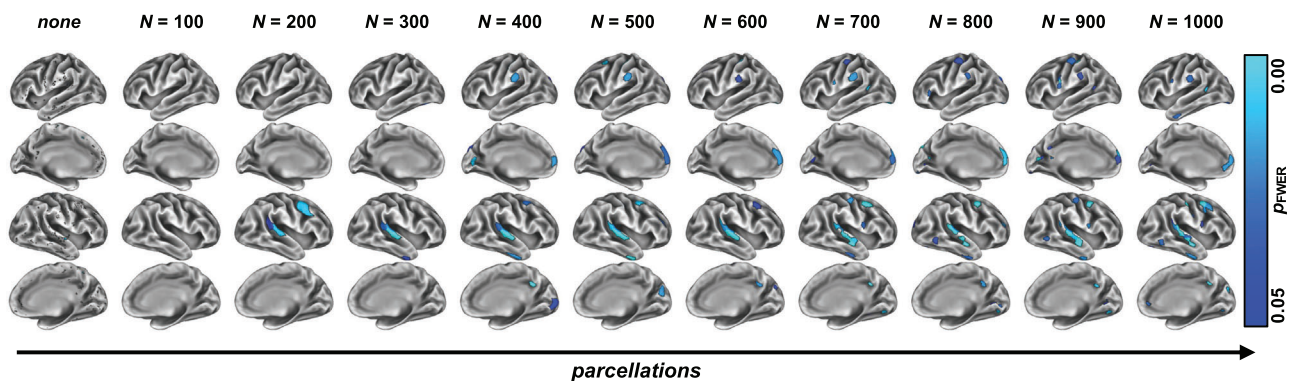
applied for MAP, from a qualitative point of view, higher resolution parcellation schemes output sharper defined individual atrophy maps and might therefore be more interesting for clinical interpretation. Another aspect which makes MAP interesting for clinical application is its potential for straightforward visualization of personalized atrophy maps (e.g., Figure 7).

#### 4.1 | Widespread gray matter atrophy in multiple sclerosis with a focus of right temporal regions

Although this study was primarily of methodological interest, we want to briefly discuss the atrophy patterns identified. In particular, the CT comparisons indicated widespread cortical atrophy in the RRMS patient group as compared to HC. This is in line with previous



**FIGURE 6** Comparing methods to assess gray matter (GM) changes for clinical relevance. We sought to assess which combination of “standard approach” strategies to calculate GM variations (i.e., cortical thickness [CT], voxel-based morphometry [VBM], CAT12 || smoothing of  $\sigma = 0/2/3/4$  mm || correction for multiple comparisons as threshold-free cluster enhancement [TFCE] or false discovery rate [FDR]) and/or the mosaic approach (MAP), that is, a method based on parcellating the cortex and calculating which of those parcels deviated from a control population for single patients, yielded highest correlations with two clinically relevant variables, namely, lesion volume (a) expanded disability status scores (EDSS) (b). We found that the MAP yielded higher correlations than any combination of the standard approaches for both lesion volume (LV) (green crosses in a) and EDSS (green crosses in b), whereas the benefits for higher resolution parcellation schemes are neglectable in terms of such quantitative clinical relevance (as opposed to qualitative relevance as demonstrated in this figure, where higher resolution parcellation scheme produces more sharply defined and thus clinically interpretable output maps). Among the standard approaches, family-wise error (FWE)-correction (red symbols) produced higher correlations versus FDR-correction (blue symbols). In terms of smoothing, we found that smoothing in general improved correlations, ever-more smoothing impaired the results (cf. moving right-ward on the subpanels). CT (circles) was observed to correlate higher with LV, whereas VBM (triangles) yielded higher correlations with EDSS, and both those metrics exceeded GM content (i.e., CAT12, squares). In general, GM variation was stronger correlated with LV as compared to EDSS scores (compare mean levels of all symbols in a vs. b). CAT12, computational anatomy toolbox, version 12 (used for calculating GM content); HC, healthy controls; RRMS, relapsing-remitting multiple sclerosis.



**FIGURE 7** Visualizing the mosaic approach (MAP) for individual patients. In addition to the MAP producing more clinically relevant maps of affected gray matter (GM) for single relapsing-remitting multiple sclerosis (RRMS) patients versus healthy control (HC) (e.g., Figure 6), this method is well-suited for visualizing individual maps of GM damage. Here, we show the results for a mildly affected 36-year old female patient (EDSS = 1.5, LV = 0.88 mL). Even for this clinically mildly affected patient, especially the higher resolution maps (i.e., moving rightwards on the figure) suggest established GM damage, focused on the right temporal and premotor cortices. EDSS, expanded disability status score; LV, lesion volume.

studies demonstrating both baseline and progressive widespread cortical atrophy in MS (Bergsland et al., 2012; Hidalgo de la Cruz et al., 2021; Narayana et al., 2012; Sailer et al., 2003; Tillema et al., 2016; Tsagkas et al., 2020). Intriguingly, recent work suggests that the topographical patterns of atrophy are nonrandom but that there is an anatomical organization principle to atrophy patterns (Steenwijk et al., 2016). Of note, the locations of WM lesions, more specifically the con-

nected cortical somata, have been found to be strongly correlated (Bussas et al., 2022). Interestingly, we found such widespread differences primarily using CT methods; cortical GM content and volume methods suggested a focus on the temporal cortex. Interestingly, temporal atrophy is related to cognitive decline in MS (Tillema et al., 2016), and—via the association to the limbic system—is affected in the early disease course (Audoin et al., 2010). This is in line with the present



study's patient population who yield relatively short disease duration ( $8.65 \text{ years} \pm 5.16$ ), which suggests that the CT parcellation method is more sensitive to detect subtle differences as opposed to the standard approaches, which might become clinically relevant later on in the disease course.

## 4.2 | The potential of MAP as a clinically relevant biomarker in MS and beyond

The main takeaway from this study is that we provide evidence that a clinically relevant biomarker for MS is particularly constructable from parcellated CT data, as realized in the MAP method, which can be (1) easily calculated, (2) applied to, and (3) interpreted for single patients. All of these three characteristics are fundamental for clinical translation of biomarker. Of note, we provide evidence that MAP—that is, referencing single subject's high-resolution parcellation scheme of the CT data and interpreting distinct patches with respect to a reference group—yields higher correlations with external clinical metrics, namely, LV and EDSS, and might therefore be a promising strategy for quantifying and interpreting GM atrophy in individual patients using MRI. These observations suggest the generalizability of the MAP method—which we have previously shown to be clinically relevant in MND and FTD (McKenna et al., 2022; Tahedl et al., 2021, 2022)—also to MS, where it might help fill the need for a personalized biomarker to appreciate single-patient cortical involvement (Amiri et al., 2018).

## 4.3 | Limits

Although the results of the current study are promising for helping the advancement of MAP a biomarker for estimating cortical disease burden, it comes with certain limits and should therefore be interpreted with caution. First of all, we want to emphasize that we are far from considering all documented strategies and correction methods for assessing GM atrophy and therefore we cannot claim to demonstrate MAP as “the ideal” biomarker. Moreover, our statistical comparisons might suffer from an overall relatively small HC reference group ( $N = 89$ ), which results in discretely distributed  $p$ -values in permutation testing, as we did, here, which might in turn hamper statistical power (Nichols & Holmes, 2002). Also note that the original MAP method suggests to further subdivide the HC reference groups into smaller age- and gender-matches subgroups (Tahedl et al., 2021), which we could not provide with the present reference group for its sample size, as outlined above. Note that in the present study, we used an internal reference group to compute MAP, that is, patient and control data were acquired at the same scanner using the same scanning parameters. However, evidence from our previous work suggests that MAP is also valid when using control data from external reference groups acquired using different scanning parameters, such as the HCP (Tahedl, 2020) or the CamCAN repository (Tahedl et al., 2021). Indeed, recent evidence further supports the validity of using external reference data to assess single subject T1w data, which might however require a criti-

cal mass of large-scale control data sets (Bethlehem et al., 2022). Larger and more variable reference groups allow for testing with less discrete distributions of  $p$ -values and hence enhance statistical power. However, latter two limits might also be interpreted as a strength of the study's results since the fact that we did find differences suggests strong effect sizes. Moreover, controlling for documented effects on GM estimation related to technical influences, such as TI or the scanner itself (Biberacher et al., 2016; Durand-Dubief et al., 2012), could not be accounted for in this data set and needs to be further investigated specifically for MAP. Importantly, the reliability of MAP needs to be further evaluated, although some longitudinal reports for this method exist, which at least provide some first evidence of validity of MAP for monitoring clinical progression in ALS and PLS (Tahedl et al., 2021, 2022). Finally, and fundamental for any MS biomarker, is that a metric purely based on cortical disease burden cannot reflect the complexity of the pathology, which—as outlined throughout this manuscript—manifests widespread throughout the CNS, including WM, deep GM, and spinal cord. Nevertheless, we provided evidence that the purely cortical biomarker MAP does transfer clinically relevant information in MS and is therefore a candidate biomarker for monitoring aspects of MS disease state/progression as well as serving as a target for clinical trials. Future studies will need to further refine and investigate MAP to help its translation into daily neuroradiological clinical routine, both for MS and other conditions with cortical involvement—including MND and FTD, for which the method was already adopted—especially in terms of longitudinal reliability and validity, to reach the critical need for a personalized biomarker reflecting single-patient cortical involvement (Amiri et al., 2018).

## 5 | CONCLUSIONS

The present study investigated different methods to assess cortical disease burden for individual MS patients. We found that the mosaic approach or “MAP”—a relatively novel biomarker based on estimating cortical disease burden using CT estimations from high-resolution parcellation schemes—yields high potential for a clinically relevant biomarker in MS, outperforming existing methods to compute cortical disease burden in single patients. Of note, MAP outputs brain maps illustrating individual cortical disease burden which can be directly interpreted in daily clinical routine.

### AUTHOR CONTRIBUTIONS

**Marlene Tahedl:** Conceptualization; formal analysis; funding acquisition; investigation; methodology; resources; validation; visualization; writing—original draft; writing—review and editing. **Tun Wiltgen:** Investigation; writing—review and editing. **Cui Ci Voon:** Writing—review and editing. **Achim Berthele:** Resources; supervision; writing—review and editing. **Jan S. Kirschke:** Data curation; funding acquisition; project administration; resources; supervision; writing—review and editing. **Bernhard Hemmer:** Data curation; funding acquisition; project administration; resources; supervision; writing—review and editing. **Mark Mühlau:** Conceptualization; data curation; methodology;

resources; supervision; writing—original draft; writing—review and editing. **Claus Zimmer**: Data curation; funding acquisition; project administration; resources; supervision; writing—review and editing. **Benedikt Wiestler**: Conceptualization; formal analysis; funding acquisition; methodology; project administration; supervision; writing—original draft; writing—review and editing.

## ACKNOWLEDGMENTS

BH is associated with DIFUTURE (Data Integration for Future Medicine) [BMBF 01ZZ1804[A-I]] and received funding for the study by the European Union's Horizon 2020 Research and Innovation Program [grant MultipleMS, EU RIA 733161] and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology [EXC 2145 SyNergy—ID 390857198].

Open access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

AB has received consulting and/or speaker fees from Alexion, Biogen, Celgene, Horizon, Novartis, Roche and Sandoz/Hexal, and his institution has received compensation for clinical trials from Alexion, Biogen, Merck, Novartis, Roche, and Sanofi Genzyme, all outside the current work. JSK has received speaker fees from Novartis. BH has served on scientific advisory boards for Novartis; he has served as DMSC member for AllergyCare, Sandoz, Polpharma, Biocon, and TG therapeutics; his institution received research grants from Roche for multiple sclerosis research. He has received honoraria for counseling (Gerson Lehrman Group). He holds part of two patents; one for the detection of antibodies against KIR4.1 in a subpopulation of patients with multiple sclerosis and one for genetic determinants of neutralizing antibodies to interferon. All conflicts are not relevant to the topic of the study.

## DISCLOSURES

AB has received consulting and/or speaker fees outside of the current work. The details are at the end of the manuscript. The other authors declare no competing interests.

## DATA AVAILABILITY STATEMENT

NA.

## ORCID

Marlene Tahedl  <https://orcid.org/0000-0003-1762-4824>

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/brb3.3327>.

## REFERENCES

- Amiri, H., de Sitter, A., Bendfeldt, K., Battaglini, M., Wheeler-Kingshott, C. A. M. G., Calabrese, M., Geurts, J. J. G., Rocca, M. A., Sastre-Garriga, J., Enzinger, C., de Stefano, N., Filippi, M., Rovira, Á., Barkhof, F., & Vrenken, H. (2018). Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. In *Neurolmage: Clinical* (Vol. 19, pp. 466–475). Elsevier Inc. <https://doi.org/10.1016/j.nicl.2018.04.023>
- Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry—The methods. *Neuroimage*, 11(6 Pt 1), 805–821. <https://doi.org/10.1006/nimg.2000.0582>
- Audoïn, B., Zaaoui, W., Reuter, F., Rico, A., Malikova, I., Confort-Gouny, S., Cozzone, P. J., Pelletier, J., & Ranjeva, J.-P. (2010). Atrophy mainly affects the limbic system and the deep grey matter at the first stage of multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 81(6), 690–695. <https://doi.org/10.1136/jnnp.2009.188748>
- Bakshi, R., Benedict, R. H., Bermel, R. A., & Jacobs, L. (2001). Regional brain atrophy is associated with physical disability in multiple sclerosis: Semiquantitative magnetic resonance imaging and relationship to clinical findings. *Journal of Neuroimaging: Official Journal of the American Society of Neuroimaging*, 11(2), 129–136. <https://doi.org/10.1111/j.1552-6569.2001.tb00022.x>
- Barkhof, F. (2002). The clinico-radiological paradox in multiple sclerosis revisited. *Current Opinion in Neurology*, 15(3), 239–245. <https://doi.org/10.1097/00019052-200206000-00003>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1), 289–300. <https://doi.org/10.2307/2346101>
- Bergsland, N., Horakova, D., Dwyer, M. G., Dolezal, O., Seidl, Z. K., Vaneckova, M., Krasensky, J., Havrdova, E., & Zivadinov, R. (2012). Subcortical and cortical gray matter atrophy in a large sample of patients with clinically isolated syndrome and early relapsing-remitting multiple sclerosis. *AJNR. American Journal of Neuroradiology*, 33(8), 1573–1578. <https://doi.org/10.3174/ajnr.A3086>
- Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., ... Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, 604, 525–533. <https://doi.org/10.1038/s41586-022-04554-y>
- Biberacher, V., Schmidt, P., Keshavan, A., Boucard, C. C., Righart, R., Sämann, P., Preibisch, C., Fröbel, D., Aly, L., Hemmer, B., Zimmer, C., Henry, R. G., & Mühlau, M. (2016). Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. *Neuroimage*, 142, 188–197. <https://doi.org/10.1016/j.neuroimage.2016.07.035>
- Bussas, M., Grahl, S., Pongratz, V., Berthele, A., Gasperi, C., Andlauer, T., Gaser, C., Kirschke, J. S., Wiestler, B., Zimmer, C., Hemmer, B., & Mühlau, M. (2022). Gray matter atrophy in relapsing-remitting multiple sclerosis is associated with white matter lesions in connecting fibers. *Multiple Sclerosis Journal*, 28(6), 900–909. <https://doi.org/10.1177/13524585211044957>
- Cash, D. M., Frost, C., Iheme, L. O., Ünay, D., Kandemir, M., Fripp, J., Salvado, O., Bourgeat, P., Reuter, M., Fischl, B., Lorenzi, M., Frisoni, G. B., Pennec, X., Pierson, R. K., Gunter, J. L., Senjem, M. L., Jack, C. R., Guizard, N., Fonov, V. S., ..., Ourselin, S. (2015). Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge. *Neuroimage*, 123, 149–164. <https://doi.org/10.1016/j.neuroimage.2015.07.087>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- de Stefano, N., Matthews, P. M., Filippi, M., Agosta, F., de Luca, M., Bartolozzi, M. L., Guidi, L., Ghezzi, A., Montanari, E., Cifelli, A., Federico, A., & Smith, S. M. (2003). Evidence of early cortical atrophy in MS: Relevance to white matter changes and disability. *Neurology*, 60(7), 1157–1162. <https://doi.org/10.1212/01.WNL.0000055926.69643.03>
- Dickie, E. W., Anticevic, A., Smith, D. E., Coalson, T. S., Manogaran, M., Calarco, N., Viviano, J. D., Glasser, M. F., Van Essen, D. C., & Voineskos, A. N. (2019). Ciftify: A framework for surface-based analysis of legacy MR acquisitions. *Neuroimage*, 197, 818–826. <https://doi.org/10.1016/j.neuroimage.2019.04.078>

- Durand-Dubief, F., Belaroussi, B., Armspach, J. P., Dufour, M., Roggerone, S., Vukusic, S., Hannoun, S., Sappey-Mariniere, D., Confavreux, C., & Cotton, F. (2012). Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: Comparison of 7 quantification techniques. *AJNR. American Journal of Neuroradiology*, 33(10), 1918–1924. <https://doi.org/10.3174/ajnr.A3107>
- Eijlers, A. J. C., Dekker, I., Steenwijk, M. D., Meijer, K. A., Hulst, H. E., Pouwels, P. J. W., Uitdehaag, B. M. J., Barkhof, F., Vrenken, H., Schoonheim, M. M., & Geurts, J. J. G. (2019). Cortical atrophy accelerates as cognitive decline worsens in multiple sclerosis. *Neurology*, 93(14), e1348–e1359. <https://doi.org/10.1212/WNL.0000000000008198>
- Engl, C., Tiemann, L., Grahl, S., Bussas, M., Schmidt, P., Pongratz, V., Berthele, A., Beer, A., Gaser, C., Kirschke, J. S., Zimmer, C., Hemmer, B., & Mühlau, M. (2020). Cognitive impairment in early MS: Contribution of white matter lesions, deep grey matter atrophy, and cortical atrophy. *Journal of Neurology*, 267(8), 2307–2318. <https://doi.org/10.1007/s00415-020-09841-0>
- Eshaghi, A., Marinescu, R. V., Young, A. L., Firth, N. C., Prados, F., Jorge Cardoso, M., Tur, C., de Angelis, F., Cawley, N., Brownlee, W. J., de Stefano, N., Laura Stromillo, M., Battaglini, M., Ruggieri, S., Gasperini, C., Filippi, M., Rocca, M. A., Rovira, A., Sastre-Garriga, J., ..., Ciccarelli, O. (2018). Progression of regional grey matter atrophy in multiple sclerosis. *Brain: A Journal of Neurology*, 141(6), 1665–1677. <https://doi.org/10.1093/brain/awy088>
- Filippi, M., & Rocca, M. A. (2010). MR imaging of gray matter involvement in multiple sclerosis: Implications for understanding disease pathophysiology and monitoring treatment efficacy. *AJNR. American Journal of Neuroradiology*, 31(7), 1171–1177. <https://doi.org/10.3174/ajnr.A1944>
- Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20), 11050–11055. <https://doi.org/10.1073/pnas.200033797>
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355. <http://www.ncbi.nlm.nih.gov/pubmed/11832223>
- Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, 23, S69–S84. <https://doi.org/10.1016/j.neuroimage.2004.07.016>
- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2), 195–207. <https://doi.org/10.1006/nimg.1998.0396>
- Gaser, C., Dahnke, R., Thompson, P. M., Kurth, F., Luders, E., & Alzheimer's Disease Neuroimaging Initiative. (2022). CAT—A computational anatomy toolbox for the analysis of structural MRI data. *BioRxiv*, [Preprint]. <https://doi.org/10.1101/2022.06.11.495736>
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4), 870–878. <https://doi.org/10.1006/nimg.2001.1037>
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Essen, D. C. V., Jenkinson, M., & Hcp, W. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*, 80, 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>
- Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Fristen, K. J., & Frackowiak, R. S. J. (2002). A voxel-based morphometric study of ageing in 465 normal adult human brains. *Biomedical Imaging V—Proceedings of the 5th IEEE EMBS International Summer School on Biomedical Imaging, SSBI 2002* (Vol. 1998, pp. 21–36). IEEE. <https://doi.org/10.1109/SSBI.2002.1233974>
- Gunter, J. L., Shiung, M. M., Manduca, A., & Jack, C. R. (2003). Methodological considerations for measuring rates of brain atrophy. *Journal of Magnetic Resonance Imaging*, 18(1), 16–24. <https://doi.org/10.1002/jmri.10325>
- Hidalgo de la Cruz, M., Valsasina, P., Gobbi, C., Gallo, A., Zecca, C., Biseco, A., Tedeschi, G., Filippi, M., & Rocca, M. A. (2021). Longitudinal cortical thinning progression differs across multiple sclerosis phenotypes and is clinically relevant: A multicentre study. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 27(6), 827–840. <https://doi.org/10.1177/1352458520940548>
- Hopfinger, J. B., Büchel, C., Holmes, A. P., & Friston, K. J. (2000). A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. *Neuroimage*, 11(4), 326–333. <https://doi.org/10.1006/nimg.2000.0549>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *Neuroimage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Jo, H. J., Lee, J.-M., Kim, J.-H., Shin, Y.-W., Kim, I.-Y., Kwon, J. S., & Kim, S. I. (2007). Spatial accuracy of fMRI activation influenced by volume- and surface-based spatial smoothing techniques. *Neuroimage*, 34(2), 550–564. <https://doi.org/10.1016/j.neuroimage.2006.09.047>
- Lassmann, H. (2018). Multiple Sclerosis Pathology. *Cold Spring Harbor Perspectives in Medicine*, 8(3), a028936. <https://doi.org/10.1101/cshperspect.a028936>
- Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S. W., & Van Essen, D. C. (2011). Informatics and data mining tools and strategies for the Human Connectome Project. *Frontiers in Neuroinformatics*, 5(June), 1–12. <https://doi.org/10.3389/fninf.2011.00004>
- McKenna, M. C., Tahedi, M., Lope, J., Chipika, R. H., Li Hi Shing, S., Doherty, M. A., Hengeveld, J. C., Vajda, A., McLaughlin, R. L., Hardiman, O., Hutchinson, S., & Bede, P. (2022). Mapping cortical disease-burden at individual-level in frontotemporal dementia: Implications for clinical care and pharmacological trials. *Brain Imaging and Behavior*, 16(3), 1196–1207. <https://doi.org/10.1007/s11682-021-00523-7>
- Mendelsohn, Z., Pemberton, H. G., Gray, J., Goodkin, O., Carrasco, F. P., Scheel, M., Nawabi, J., & Barkhof, F. (2023). Commercial volumetric MRI reporting tools in multiple sclerosis: A systematic review of the evidence. *Neuroradiology*, 65(1), 5–24. <https://doi.org/10.1007/s00234-022-03074-w>
- Mollison, D., Sellar, R., Bastin, M., Mollison, D., Chandran, S., Wardlaw, J., & Connick, P. (2017). The clinico-radiological paradox of cognitive function and MRI burden of white matter lesions in people with multiple sclerosis: A systematic review and meta-analysis. *PLoS ONE*, 12(5), e0177727. <https://doi.org/10.1371/journal.pone.0177727>
- Mortazavi, D., Kouzani, A. Z., & Soltanian-Zadeh, H. (2012). Segmentation of multiple sclerosis lesions in MR images: A review. *Neuroradiology*, 54(4), 299–320. <https://doi.org/10.1007/s00234-011-0886-7>
- Narayana, P. A., Govindarajan, K. A., Goel, P., Datta, S., Lincoln, J. A., Cofield, S. S., Cutter, G. R., Lublin, F. D., Wolinsky, J. S., & MRI Analysis Center at Houston, The CombiRx Investigators Group. (2012). Regional cortical thickness in relapsing remitting multiple sclerosis: A multi-center study. *NeuroImage. Clinical*, 2, 120–131. <https://doi.org/10.1016/j.nicl.2012.11.009>
- Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1), 1–25. <https://doi.org/10.1002/hbm.1058>
- Popescu, V., Agosta, F., Hulst, H. E., Sluimer, I. C., Knol, D. L., Sormani, M. P., Enzinger, C., Ropele, S., Alonso, J., Sastre-Garriga, J., Rovira, A., Montalban, X., Bodini, B., Ciccarelli, O., Khaleeli, Z., Chard, D. T., Matthews, L., Palace, J., Giorgio, A., ... Vrenken, H. (2013). Brain atrophy and lesion load predict long term disability in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 84(10), 1082–1091. <https://doi.org/10.1136/jnnp-2012-304094>

- Popescu, V., Schoonheim, M. M., Versteeg, A., Chaturvedi, N., Jonker, M., Xavier de Menezes, R., Gallindo Garre, F., Uitdehaag, B. M. J., Barkhof, F., & Vrenken, H. (2016). Grey matter atrophy in multiple sclerosis: Clinical interpretation depends on choice of analysis method. *PLoS ONE*, 11(1), e0143942. <https://doi.org/10.1371/journal.pone.0143942>
- Sailer, M., Fischl, B., Salat, D., Tempelmann, C., Schönfeld, M. A., Busa, E., Bodammer, N., Heinze, H.-J., & Dale, A. (2003). Focal thinning of the cerebral cortex in multiple sclerosis. *Brain: A Journal of Neurology*, 126(Pt 8), 1734–1744. <https://doi.org/10.1093/brain/awg175>
- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9), 3095–3114. <https://doi.org/10.1093/cercor/bhx179>
- Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förchler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., Hemmer, B., & Mühlau, M. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage*, 59(4), 3774–3783. <https://doi.org/10.1016/j.neuroimage.2011.11.032>
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751–754. <https://doi.org/10.1093/biomet/73.3.751>
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1), 83–98. <https://doi.org/10.1016/j.neuroimage.2008.03.061>
- Sormani, M. P., Arnold, D. L., & de Stefano, N. (2014). Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Annals of Neurology*, 75(1), 43–49. <https://doi.org/10.1002/ana.24018>
- Steenwijk, M. D., Geurts, J. J. G., Daams, M., Tijms, B. M., Wink, A. M., Balk, L. J., Tewarie, P. K., Uitdehaag, B. M. J., Barkhof, F., Vrenken, H., & Pouwels, P. J. W. (2016). Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain*, 139(1), 115–126. <https://doi.org/10.1093/brain/awv337>
- Tahedl, M. (2020). Towards individualized cortical thickness assessment for clinical routine. *Journal of Translational Medicine*, 18(1), 151. <https://doi.org/10.1186/s12967-020-02317-9>
- Tahedl, M., Chipika, R. H., Lope, J., Li Hi Shing, S., Hardiman, O., & Bede, P. (2021). Cortical progression patterns in individual ALS patients across multiple timepoints: A mosaic-based approach for clinical use. *Journal of Neurology*, 268(5), 1913–1926. <https://doi.org/10.1007/s00415-020-10368-7>
- Tahedl, M., Li Hi Shing, S., Finegan, E., Chipika, R. H., Lope, J., Hardiman, O., & Bede, P. (2022). Propagation patterns in motor neuron diseases: Individual and phenotype-associated disease-burden trajectories across the UMN-LMN spectrum of MNDs. *Neurobiology of Aging*, 109, 78–87. <https://doi.org/10.1016/j.neurobiolaging.2021.04.031>
- RStudio Team. (2022). *RStudio: Integrated development environment for R*. RStudio PBC. <http://www.rstudio.com/>
- Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., ... Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2), 162–173. [https://doi.org/10.1016/S1474-4422\(17\)30470-2](https://doi.org/10.1016/S1474-4422(17)30470-2)
- Tillema, J. M., Hulst, H. E., Rocca, M. A., Vrenken, H., Steenwijk, M. D., Damjanovic, D., Enzinger, C., Ropele, S., Tedeschi, G., Gallo, A., Ciccarelli, O., Rovira, A., Montalban, X., de Stefano, N., Stromillo, M. L., Filippi, M., & Barkhof, F., & MAGNIMS Study Group. (2016). Regional cortical thinning in multiple sclerosis and its relation with cognitive impairment: A multicenter study. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 22(7), 901–909. <https://doi.org/10.1177/1352458515607650>
- Todea, R.-A., Lu, P.-J., Fartaria, M. J., Bonnier, G., Du Pasquier, R., Krueger, G., Bach Cuadra, M., Psychogios, M. N., Kappos, L., Kuhle, J., & Granziera, C. (2020). Evolution of cortical and white matter lesion load in early-stage multiple sclerosis: Correlation with neuroaxonal damage and clinical changes. *Frontiers in Neurology*, 11, 973. <https://doi.org/10.3389/fneur.2020.00973>
- Tsagkas, C., Chakravarty, M. M., Gaetano, L., Naegelin, Y., Amann, M., Parmar, K., Papadopoulou, A., Wufer, J., Kappos, L., Sprenger, T., & Magon, S. (2020). Longitudinal patterns of cortical thinning in multiple sclerosis. *Human Brain Mapping*, 41(8), 2198–2215. <https://doi.org/10.1002/hbm.24940>
- Tsouki, F., & Williams, A. (2021). Multifaceted involvement of microglia in gray matter pathology in multiple sclerosis. *Stem Cells*, 39(8), 993–1007. <https://doi.org/10.1002/stem.3374>
- Vrenken, H., & Geurts, J. J. G. (2007). Gray and normal-appearing white matter in multiple sclerosis: An MRI perspective. *Expert Review of Neurotherapeutics*, 7(3), 271–279. <https://doi.org/10.1586/14737175.7.3.271>

**How to cite this article:** Tahedl, M., Wiltgen, T., Voon, C. C., Berthele, A., Kirschke, J. S., Hemmer, B., Mühlau, M., Zimmer, C., & Wiestler, B. (2023). Benefits of a mosaic approach for assessing cortical atrophy in individual multiple sclerosis patients. *Brain and Behavior*, 13, e3327. <https://doi.org/10.1002/brb3.3327>