Regional Brain Volumetry and Brain Function in Severely Brain-Injured Patients

Jitka Annen, MSc,[1,2] Gianluca Frasso, Ph.D.,[3] Julia Sophia Crone, Ph.D.,[4]
Lizette Heine, Ph.D.,[1,5] Carol Di Perri, M.D., Ph.D. (ORCID)[1,2,6]
Charlotte Martial, MSc,[1,2] Helena Cassol, MSc,[1,2] Athena Demertzi, Ph.D.,[1,7]
Lionel Naccache, M.D., Ph.D.,[7] Steven Laureys, M.D., Ph.D.,[1,2]
and Coma Science Group Collaborators

Objective: The relationship between residual brain tissue in patients with disorders of consciousness (DOC) and the clinical condition is unclear. This observational study aimed to quantify gray (GM) and white matter (WM) atrophy in states of (altered) consciousness.

Methods: Structural T1-weighted magnetic resonance images were processed for 102 severely brain-injured and 52 healthy subjects. Regional brain volume was quantified for 158 (sub)cortical regions using Freesurfer. The relationship between regional brain volume and clinical characteristics of patients with DOC and conscious brain-injured patients was assessed using a linear mixed-effects model. Classification of patients with unresponsive wakefulness syndrome (UWS) and minimally conscious state (MCS) using regional volumetric information was performed and compared to classification using cerebral glucose uptake from fluorodeoxyglucose positron emission tomography. For validation, the T1-based classifier was tested on independent datasets.

Results: Patients were characterized by smaller regional brain volumes than healthy subjects. Atrophy occurred faster in UWS compared to MCS (GM) and conscious (GM and WM) patients. Classification was successful (misclassification with leave-one-out cross-validation between 2% and 13%) and generalized to the independent data set with an area under the receiver operator curve of 79% (95% confidence interval [CI; 67–91.5]) for GM and 70% (95% CI [55.6–85.4]) for WM.

Interpretation: Brain volumetry at the single-subject level reveals that regions in the default mode network and sub-cortical gray matter regions, as well as white matter regions involved in long range connectivity, are most important to distinguish levels of consciousness. Our findings suggest that changes of brain structure provide information in addition to the assessment of functional neuroimaging and thus should be evaluated as well.

ANN NEUROL 2018;00:000–000

After severe brain damage, postcomatose survivors might suffer from prolonged disorders of consciousness (DOC). Deficits attributed to brain damage are assessed during clinical examination,[1] for which standardized behavioral assessment is still the most sensitive and reliable tool.[2] To this end, clinicians use behavioral scales, preferably the Coma Recovery Scale-Revised (CRS-R[3]),

administered several times.[4] Patients with unresponsive wakefulness syndrome (UWS) are fully unaware of themselves and their environment.[5] Patients in the minimally conscious state (MCS) show behavioral signs of consciousness in a fluctuating manner, which makes them difficult to detect even for experienced clinicians.[6] Patients who recover the ability to functionally use objects

or communicate have emerged from the minimally conscious state (EMCS)[6] and are no longer considered in a DOC. Locked-in syndrome (LIS) patients suffer from apraxia and quadriplegia, limiting their motor function and causing misdiagnosis even though they are fully conscious.[7] The diagnostic process, that is, the assessment of the level of consciousness depending on behavioral observation, is inherently affected by a certain level of uncertainty beyond control of the clinician attributed to, for instance, the patient's sensory and motor disabilities.[8] This study aims to identify objective measures derived from conventional magnetic resonance imaging (MRI), which is widely available, and thus may increase the understanding of the relationship between brain structure and consciousness to guide future diagnostic assessment. To date, the most sensitive neuroimaging technique to detect brain function related to residual consciousness is positron emission tomography (PET) to measure cerebral glucose uptake.[9] Consistency between the clinical and PET-based diagnosis is present in 85% of the cases.[9,10]

Conventional T1-weighted MRI is often evaluated by neuroradiologists or used for the analysis of other MRI sequences such as functional MRI. However, the literature regarding the use of T1-weighted MRI in chronic DOC patients is rather limited. Damage in the thalamus and in the basal ganglia have been related to awareness and wakefulness respectively.[11] A voxel-based morphometry study, in which every subject's brain is normalized to a template and atrophy is quantified voxelwise, indicates possible gray matter (GM) differences between MCS and UWS patient groups.[12] These studies emphasize the potential profit of conventional MRI for objective assessment of DOC patients, but indicate some limitations, for example, that the parcellation procedures are not automated, the data are aggregated on the group level, or they focus exclusively on subcortical regions.

To improve understanding of the anatomical pathophysiology and the resulting DOC, we aimed at exploring GM and white matter (WM) volumetry at a single-subject level. The fully automated approach adopted here considers all GM and WM and can be performed on individual subjects even with small brain deformations. Because the utilized methods rely on the single subject's anatomy rather than on normalization approaches, which are problematic in cases of severe brain damage due to underestimation of the lesion volume,[13] a region of interest (ROI) based analysis is conducted instead of a voxel-based one. Clinical assessment of consciousness was done with repeated assessments of the CRS-R[3] and was related to the regional volume of 84 GM and 74 WM cortical and subcortical regions in patients, that is, UWS (n = 30), MCS (n = 49), as well as conscious brain-injured patients (n = 23

including EMCS [n = 19] and LIS [n = 4]) and a group of healthy subjects (n = 52). In this study, we (1) tested whether regional volume differs across groups, (2) described the influence of disease duration and other variables (ie, age, sex, time since onset, and etiology) on brain volume at the group-level, (3) assessed the possibility to discriminate UWS and MCS patients using GM and WM regional volume, and compared the classification performance to the PET-based diagnosis, (4) identified the regions with highest discriminatory power between UWS and MCS patients, and (5) evaluated the reproducibility of the classifier's performance on independent patient data sets of collaborating centers to quantify the generalizability of the classifier.

Patients and Methods

Population

MRI data of 290 patients (Fig 1A) and 52 healthy subjects were acquired from November 2009 until January 2017 at the University Hospital of Liège, Belgium. For the independent data sets, we included 38 patients and 15 healthy subjects scanned at the Neuroradiology Department of the Paris Pitié Salpêtirère (France), and 42 patients and 28 healthy subjects scanned at the Christian Doppler Klinik in Salzburg (Austria). Even though the source of brain damage is often obvious when performing quality checks during the image analysis procedure, the investigators were blinded to the patient's diagnosis. Patients were included if they presented one of the following clinical entities after a period of coma: UWS, MCS, EMCS, or LIS. Because both EMCS and LIS patients are not considered to suffer from DOC, these patients were grouped in one group (conscious). Patients were excluded for being underage, in an acute stage after brain injury (ie, scanned < 28 days post-injury), ambiguous clinical diagnosis, having metal implants not compatible with MRI scanning, and presenting brain lesions affecting more than two thirds of one hemisphere. Etiology was categorized as traumatic brain injury (TBI), or nontraumatic brain injury (ie, anoxia, anoxia plus TBI, metabolic cause, infection, and hemorrhage). For the Liège data set, the diagnosis was determined by multiple evaluations (four to seven) of the CRS-R[3] (and additional neuropsychological testing for EMCS and LIS patients) within a week of the MRI scan. For the Paris and Salzburg data set, patients were assessed at least twice with the CRS-R. The patient's outcome was assessed 6 months after hospitalization using the extended version of the Glasgow Outcome Scale (GOS-E) outcome scale.[14] This scale is the most common outcome measure after brain injury with outcomes ranging from severe disability (unconscious) to good recovery and independency. Even if the scale was initially developed for patients after a TBI, it is used for patients with any kind of brain injury.[9] Healthy subjects and legal guardians of patients have given their written informed consent according to the Declaration of Helsinki. The study was approved by the ethics committee of the University Hospital of Liège. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines were followed thoroughly.[15]
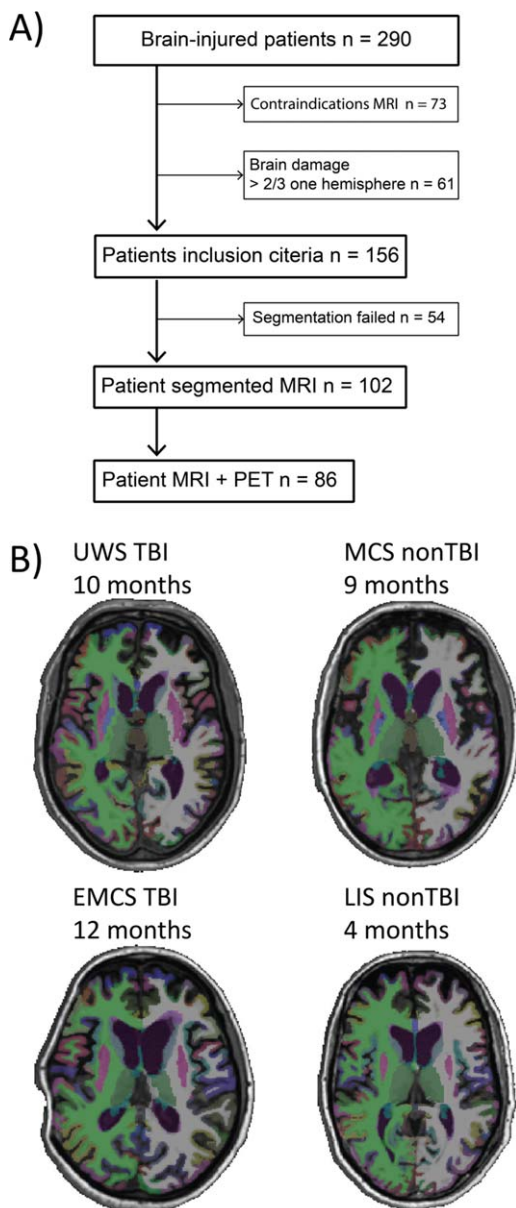
FIGURE 1: (A) Flowchart and inclusion criteria. Data in this study were collected using stringent inclusion criteria on the MRI quality and brain injury. Of the 290 adult and nonacute (ie, > 28 days postinjury), 102 could be included in the study and 86 patients had a PET available. (B) Freesurfer segmentation for a representative subject of each patient sample. Shown are four axial T1-weighted slices, with labeled parcellated gray matter, subcortical structures, and white matter according to the Desikan-Killiany atlas. UWS = unresponsive wakefulness syndrome; MCS = minimally conscious state; Conscious = represented by patients in LIS (Locked-in syndrome) and EMCS (emergence from minimally conscious state).

Data Acquisition

For the Liège data set, structural MRI data were acquired using a 3 Tesla (T) scanner (Siemens Tim Trio; Siemens Medical Solutions, Erlangen, Germany) with 120 T1-weighted slices (repetition time [TR] = 2,300 ms, echo time [TE] = 2.47 ms, flip angle 9 degrees, voxel size = $1 \times 1 \times 1.2$ mm^3, field of view [FOV] = 256 mm^2). Data from sedated and nonsedated

patients are included given the nature of the study. Fluorodeoxyglucose (^{18}F-FDG) PET were acquired on a Gemini TF PET-CT scanner (Philips Medical Systems, Best, The Netherlands) for 86 patients of the Liège data set.

The Salzburg data set was acquired on the same scanner as the Liège data (TR = 2,300 ms, TE = 2.91 ms, flip angle 9 degrees, voxel size = $1 \times 1 \times 1.2$ mm^3, FOV = 256 mm^2). The Paris data set was acquired on a 3T scanner (Signa HDxt GE Medical Systems; GE Healthcare, Little Chalfont, UK) using different protocols (12 scans with: TR = 7,636 ms, TE = 3.18 ms, flip angle, 15 degrees, voxel size = $0.49 \times 0.49 \times 1.2$ mm^3, FOV = 256 mm^2; 12 scans with the same parameters except the TR = 716 and the TE = 3.096; one scan with TR = 9,492 ms, TE = 3.672 ms, flip angle 15 degrees, voxel size = $0.49 \times 0.49 \times 1.4$ mm^3, FOV = 256 mm^2; one scan with TR = 1,800 ms, TE = 2.35 ms, flip angle 15 degrees, voxel size = $0.42 \times 0.42 \times 0.85$ mm^3, FOV = 256 mm^2; one scan with TR = 918 ms, TE = 3.584 ms, flip angle 15 degrees, voxel size = $0.49 \times 0.49 \times 0.69$ mm^3, FOV = 256 mm^2).

Data Processing

Tissue segmentation and (sub)cortical reconstruction were performed using the image analysis software Freesurfer (version 5.3.0). In short, we normalized scan intensity artifacts,[16] removed nonbrain tissue, applied an automated Talairach transformation, segmented subcortical WM and GM,[17] tessellated the GM-WM boundary,[18] and placed tissue boundaries on voxels with the biggest intensity shift.[19] The individuals' cortical model was registered to a spherical atlas based on cortical folding patterns, in order to match the cortical geometry.[20] The cortical surface was then parcellated based on the individuals' gyri and sulci pattern using the Desikan Killiany atlas.[15] The accuracy of skull stripping, segmentation, and cortical reconstruction was visually inspected for each subject, and in case of inaccuracies during any of these steps, the subject was excluded. An example of the resulting segmentation for a patient from each diagnosis can be found in Figure 1B.

Regional volumes were normalized using the (estimated) total intracranial volume (ICV) as presented elsewhere,[21,22] which has been validated against manual segmentation of ICV.[22] Separate group mean ICV were used for the healthy and patient samples from each acquisition site. A total of 84 cortical and subcortical GM regions and 74 WM regions were selected for further analysis (for description of the regions, see a previous work[23]). When required by the linear model assumptions, the logarithm of the volume in cm^3 was considered.

PET data were analyzed as described elsewhere.[9]

Statistical Analysis

Statistical analysis was performed using R software (R Foundation for Statistical Computing, Vienna, Austria).[24] χ^2 tests were used to evaluate differences in distribution of sex in the patient groups and healthy subjects, and differences in the distribution of etiology within the patient groups. Equality of mean age within the patient groups and healthy subjects was assessed

with a Welch two-sample t test. An analysis of variance (ANOVA) model was utilized to evaluate whether mean days between the onset and scan time were equal within each patient group. We tested for evidence of a hemispheric laterality effect on GM and WM volume using a two-way ANOVA with diagnosis, hemispheric laterality, and the interaction between the two as covariates. Given the results of our test on hemispheric laterality, left and right hemispheric analogous volumes were averaged. To test for equality of regional volume sizes between the two scanners, we used a Welch two sample t test on the healthy subjects from the three centers.

We tested for group differences in regional GM and WM volume between healthy and brain-injured subjects using multivariate analysis of covariance with sex and age as covariates. Pillai's trace statistics were used to test the significance of the covariates on overall regional volumes. The group differences in individual regional volume are reported by their group mean and 95% simultaneous confidence intervals (CIs), which were Bonferroni corrected (α for GM: $1-0.05/42 = 0.998$; WM: $1-0.05/40 = 0.998$).

A linear mixed-effects model (R package: lme4) was utilized to assess the effect of different demographic features on GM and WM atrophy within the patient population. In particular, we model the logarithm* of the volume in cm^3. Fixed-effects terms were included for sex, etiology, age, diagnosis, and years since onset, and the interactions between the latter two. We included random intercepts to model subject and regional effects, together with a random slope for the years since onset. Because the regional volumes appear more volatile for small values of the days since onset, we modeled the variance of the regression errors as proportional to the inverse of the value of the covariate. This allows to avoid heteroscedasticity effects. The following equation describes the linear regression model (the column symbol indicates variable interaction):

$$\log(Y_{ij}) = \beta_0 + \beta_1 YearsOnset_{ij} + \beta_2 Diagnosis_{ij}$$
$$+ \beta_3 YearsOnset_{ij} : Diagnosis_{ij} + \beta_4 Age_{ij} + \beta_5 Etiology_{ij}$$
$$+ \beta_6 Gender_{ij} + v_{0j} + b_{0i} + b_{1i} YearsOnset_{ij} + \varepsilon_{ij}$$
$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 \Omega), \text{ with } \Omega = diag(1/YearsOnset_{ij})$$
$$v_{0j} \sim N(0, \tau_0^2), \ b_{ki} \sim N(0, \psi_k^2), \ Cov(b_{ki}, b_{k'i}) = \psi_{kk'}$$

where $log(Y_{ij})$ is the volume of the i th brain region (with $i = 1, \ldots, M$) observed for the j th subject (with $j = 1, \ldots, n_i$), β_p for $p = 0, \ldots, 6$ are the fixed-effect coefficients (0 indicates the intercept), v_{0j} and b_{0i} indicate the random intercepts for subject and region, respectively, and b_{1i} indicates the random slope. Please note that we do not make any assumptions about the structure of the random effects variance-covariance matrix except that the effects for region i are independent from the ones of any other region i'.

The ability of regional volume to discriminate UWS and MCS patients was assessed through a leave-one-out boosting

*Please note that the estimated regression parameters of the log-linear model can be interpreted as percentage changes for small values of the ith coefficient, $(\exp(\beta_i) \approx 1 + \beta_i)$.

classification trees procedure (adaboost[25]). The boosting classifiers were built by using 100 stumps (ie, minimal classification trees with two terminal nodes). This method performs an intrinsic (soft) feature selection by providing a measure of regional importance alongside the classification task. The prediction accuracy for new cases was evaluated using a bootstrap leave-one-out cross-validation procedure. This method does not rely on any distributional assumption and achieves accurate error estimation in small samples,[26] thus making it an appropriate tool for our study (because we have only 49 MCS and 30 UWS patients). We generated 500 bootstrap samples with at least 3 UWS and 3 MCS patients.

The relevance of each regional brain volume (feature) was assessed by looking at the (average) classifier's importance parameter computed for each region (as formulated in a previous work[25]). This parameter identifies the influence on the classification task of a single feature as compared to the others. If all the regions would have the same relevance, the importance would be equal to 2.33% (100 divided by 43 ROIs) for GM and 2.63% (100 divided by 38 ROIs) for WM. For each subject, the classification probabilities were saved and used to select the most advantageous probability threshold.

To evaluate the diagnostic ability of the binary classifier, we produced receiver operating characteristic (ROC) curves that display the classifiers' sensitivity and specificity for various discrimination thresholds. The area under the ROC curve (AUROC), a measure varying between 0 and 1 with 0.5 reflecting a no-information rate, can be interpreted as the percentage of correctly classified randomly drawn pairs with 1 subject from the UWS and MCS groups. Additionally, the results are compared to the diagnosis as established with PET, the neuroimaging technique with highest diagnostic precision.[9] Subjects from the test data were classified by the same adaboost procedure with 100 iterations, using the probability threshold set as in the training data set. Differences in sex and etiology between the test and training data sets were assessed with a χ^2 test, and difference in mean age and year since onset were assessed with a Welch two sample t test.

Results

Of the 290 patients scanned, 156 met the inclusion criteria (Fig 1A) of the study for which segmentation of T1 MRI was attempted. The final sample consists of 102 patients (23 conscious [ie, 4 LIS, 19 EMCS], 49 MCS, and 30 UWS). GOS-E outcomes[14] and subject-specific information are reported in Supplementary Table 1. Fifty-four patients were excluded because of failure during the segmentation procedure (either the software could not come to a solution, or the software excluded brain tissue; see Supplementary Table 2 for detailed information per subject). No group differences between patient groups have been observed for etiology and mean years since onset (Table and Supplementary Table 3).

TABLE 1. Group Means (and Standard Deviation If Applicable) for Age, Sex, Etiology, and Years Since Onset for Healthy Subjects and Patients From Liège and the Independent Centers (Paris and Salzburg)

Group	Diagnosis	Mean age (SD)	Gender (M/F/NA)	Etiology (Anoxia, Mix, TBI, Other)	Years since onset (SD)
Healthy Liège, n = 52	NA	39.9(16.1)	25/21/6	NA	NA
Healthy independent centers, n = 43	NA	44.1(13.0)	18/25/0	NA	NA
Patient Liège, n=102	UWS, n = 30	48.1(16.5)	20/10/0	16,1,11,2	1.9(4.1)
	MCS, n = 49	36.7(13.0)	33/16/0	11,6,27,5	2.8(3.1)
	Conscious, n = 23	39.3(15.1)	16/7/0	9,0,13,1	3.4(5.3)
Patient independent centers, n = 55	UWS, n = 27	42.6(15.1)	19/8/0	17,1,4,4	0.5(1.0)
	MCS, n = 28	45.4(15.3)	20/8/0	10,0,6,11	0.7(1.4)

TBI = traumatic brain injury; Mix = TBI + anoxia.

The 52 healthy subjects did not differ from the patients in sex or mean age (Table and Supplementary Table 4).

The PET protocol was completed for 21 conscious patients (20 with a PET diagnosis of MCS and 1 of UWS), 45 MCS patients (40 with a PET diagnosis of MCS and 5 of UWS), and 22 UWS patients (6 with a PET diagnosis of MCS and 16 of UWS).

No effect of hemispheric laterality or an interaction between clinical state and hemispheric laterality was observed on GM or WM volume (Supplementary Table 5). Therefore, subsequent analysis was conducted on the average regional volumes of the left and right hemisphere. We did not find evidence for a difference in mean regional GM (t(864.96) = 1.941, $p = 0.053$, 95% CI [−0.001, 0.114]) and WM (t(822.02) = 0.444, $p = 0.657$, 95% CI [−326.8, 517.7]) volume within healthy subjects between the two different scanners.

Healthy Subjects Versus Brain-Injured Patients

Overall regional volume in brain-injured patients (conscious patients, MCS, and UWS) was reduced in GM and WM compared to the healthy population, and age but not sex negatively affects GM and WM volume (for Pillai statistics, see Supplementary Table 6). Looking at the difference between healthy and brain-injured subjects at the single-region level, evidence for a difference exists in all GM ROIs, except the frontal and temporal pole (Fig 2A). For the WM, there is a difference in the whole corpus callosum (ie, anterior, mid-anterior, central, mid-posterior, and posterior parts), the pre- and postcentral areas, the caudal middle frontal, pars opercularis, the

para-hippocampus, the pericalcarine, the fusiform, and lingual areas (Fig 2B).

Effect of Demographic Factors on Atrophy Within Brain-Injured Patients

We found evidence of a bigger regional brain volume in conscious patients compared to UWS patients in both GM and WM. WM volume is higher for male than for female patients. Our results highlight that a longer time postinjury (in years) is related to smaller regional volumes, regardless of diagnosis. The model estimates of the fixed effects, confidence intervals, and significance levels can be found in Figure 3A for the GM and Figure 3B for the WM analysis. In addition, we found a significant interaction effect between diagnosis and time since onset, revealing that unconscious UWS patients lose brain matter faster than MCS (GM) and conscious (GM and WM) patients (see Fig 3C,D for GM and WM analysis, respectively).

Classifying MCS and UWS Patients

Classification of DOC patients (UWS and MCS) with leave-one-out cross-validation on the T1-based GM measures had an AUROC of 96% (95% CI [92.4–99.6]), with a sensitivity of 0.92 (95% CI [0.82, 0.98]) and a specificity of 0.84 (95% CI [0.69, 0.95]). For the WM-based classification, we obtained an AUROC of 97.6% (95% CI [94.8–100]), with a sensitivity of 0.90 (95% CI [0.81, 0.97]), and a specificity of 0.96 (95% CI [0.83, 1.00]; Fig 4). This is comparable to the results of the classification based on PET glucose uptake, for which we found a sensitivity of 0.89 (95%
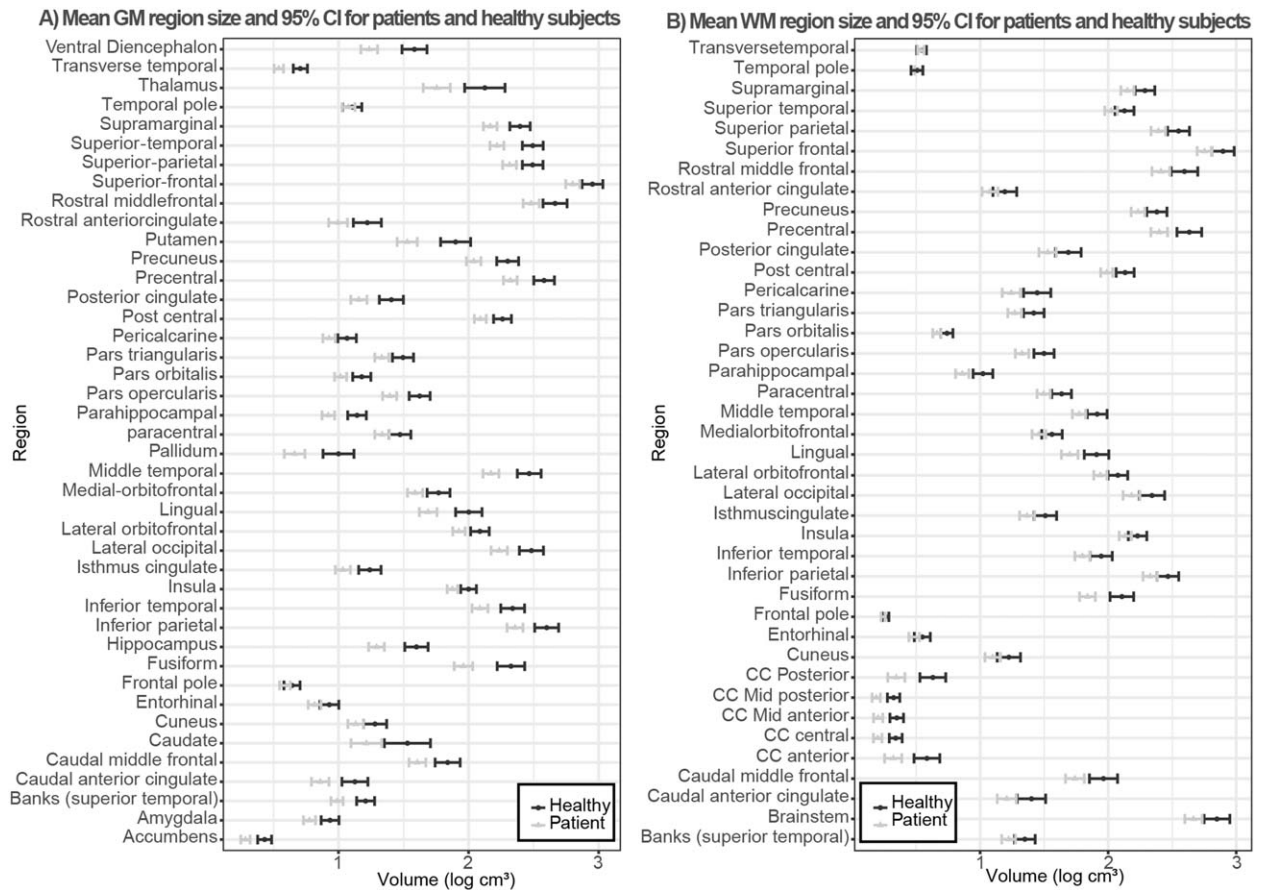
FIGURE 2: Estimates and Bonferroni corrected 95% confidence interval for regional gray (A) and white (B) matter region size for patients and healthy subjects.

CI [0.78, 0.96]) and a specificity of 0.73 (95% CI [0.53, 0.89]).

When comparing the neuroimaging diagnosis (T1 and PET) to the CRS-R diagnosis, we used only the best result of either neuroimaging modality (eg, if a patient is clinically MCS and either one of the neuroimaging diagnosis is MCS, it is considered a correct classification, but if a patient is clinically diagnosed as UWS and either one of the neuroimaging diagnosis is MCS, it is considered a false positive). There was a congruence between the CRS-R and neuroimaging for the GM in 64% of the UWS patients and 98% of the MCS patients, and for the WM in 41% of the UWS and 96% of the MCS patients (Supplementary Table 7). In only 1 and 2 MCS patients (for GM and WM, respectively), neither the T1 nor PET-based diagnosis was able to detect objective markers for consciousness while behavioral signs of MCS were observed. In 36% and 59% of the clinically diagnosed UWS patients, the T1 (for GM and WM, respectively) and/or PET were indicative for a state of MCS.

Our results furthermore show highly relevant regions for classification of UWS and MCS patients (importance level of roughly 2 times the chance level; Fig

5A,B for GM and WM, respectively). In particular, we found that the GM volume of the paracentral (7.34%), para-hippocampal (7.10%), thalamus (4.73%), caudate (4.50%), inferior parietal (4.14%), entorhinal (4.03%), and medial orbitofrontal (4.00%) cortex were key features in discriminating between patient groups. The most influencing WM regions include the rostral anterior cingulate (8.51%), banks of the superior temporal sulcus (7.85%), anterior part of the corpus callosum (6.60%), brain stem (6.45%), isthmus cingulate (6.29%), pars triangularis (6.09%), caudal anterior cingulate (5.89%), medial orbitofrontal (5.63%), para-hippocampal (5.35%), and inferior temporal (4.26%).

To quantify the generalizability of the proposed classifier, we classified an independent test data set from two different centers. Preprocessing was successful for 27 out of 38 patients from Paris, and 28 out of 42 patients from Salzburg. In total, 27 MCS and 28 UWS patients have been included in the classification protocol (Supplementary Table 1 for subject-specific demographic information, Supplementary Table 2 for subject-specific information of subjects where preprocessing failed). No difference in sex and age was found between the training
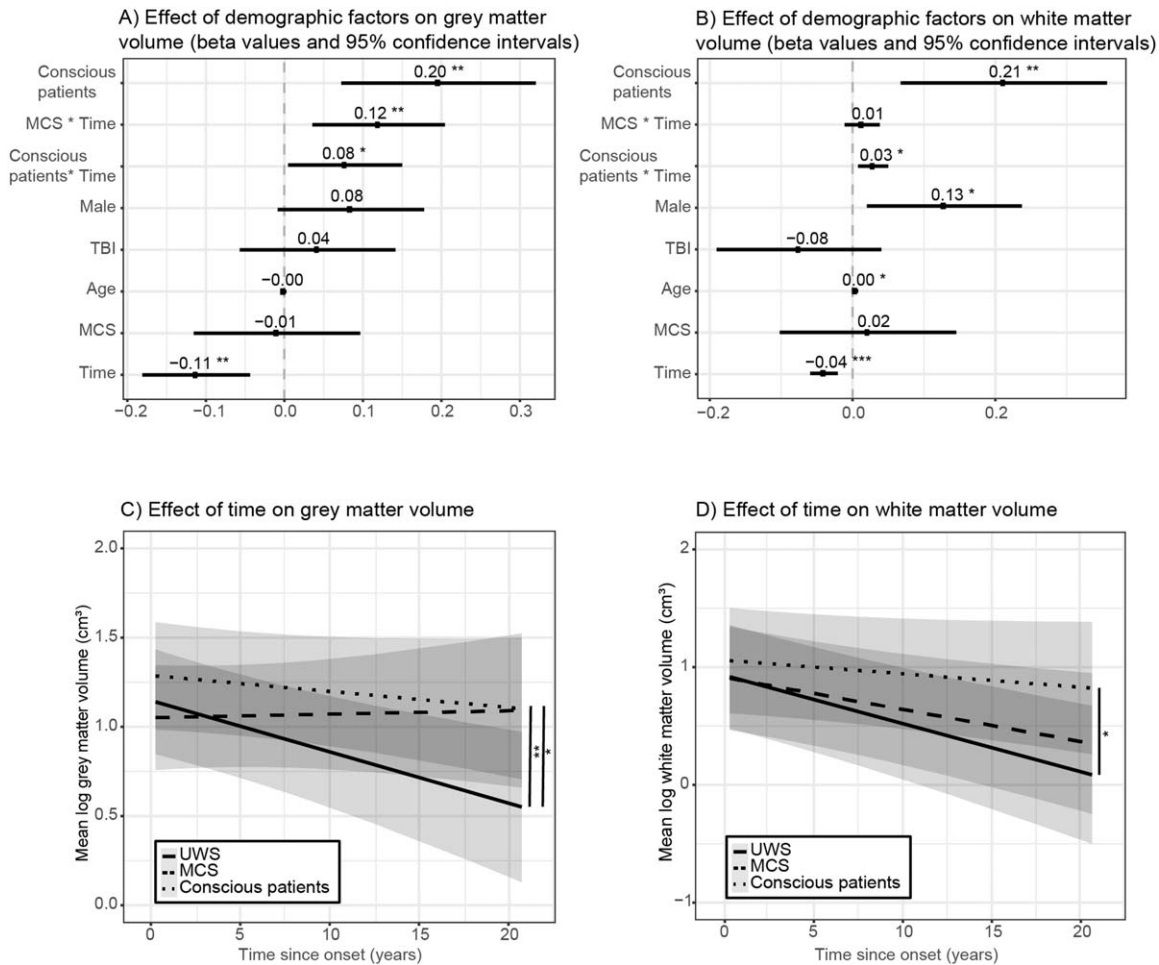
FIGURE 3: (A,C) Estimates and standard deviations of the fixed effects for the linear mixed-effects model on regional volumes, in gray (A) and white (B) matter respectively. Please note that the presented regression parameters of the log-linear model can be interpreted as expected percentage change (because $exp(\beta) \approx 1+\beta$ for small β). For example, on average, GM ROIs within the MCS patient group are 5% bigger than in UWS patients (because exp (0.05) = 1.051). MCS and conscious patients are compared to UWS patients. Sex presents how males differ from females, and for etiology is presented how TBI patients differ from non-TBI (eg, anoxia, hemorrhage). Time since onset (referred to as time) and age should be interpreted as the regional volume change per year. $*p<0.05$; $**p<0.01$; $***p<0.001$. (B,D) Estimated mean regional volume decrease (and 95% confidence interval) with increasing years since onset for each patient group in gray (C) and white (D) matter. UWS = unresponsive wakefulness syndrome, MCS = minimally conscious state, Conscious = represented by patients in LIS (Locked-in syndrome) and EMCS (emergence from minimally conscious state).

and test data (Table and Supplementary Table 4). Significant differences between the two data sets were found for time since onset and etiology (Table and Supplementary Table 3). For the GM classification, we obtained an AUROC of 79.2% (95% CI [67–91.5]), with a sensitivity of 0.85 (95% CI [0.69, 0.95]) and a specificity of 0.71 (95% CI [0.54, 0.87]). The WM classification is slightly less successful (AUROC 70.5%; 95% CI [55.6–85.4]), with a sensitivity of 0.77 (95% CI [0.61, 0.93]) and a specificity of 0.62 (95% CI [0.45, 0.80]; Fig 6). This led to a correct classification of 23 of 27 MCS patients and 20 of 28 UWS patients using GM volume. Using WM regions, we obtained a correct classification for 21 of 27 MCS patients and 17 of 27 UWS patients (Supplementary Table 8).

Discussion

In this study, we aimed at quantifying regional cortical and subcortical brain atrophy in severely brain-injured patients using a fully automated whole-brain segmentation method. As expected, brain-injured patients have reduced regional volume compared to healthy subjects in almost all brain regions. A handful of the smallest brain regions did not show any difference, possibly attributed to the limited absolute atrophy because of their small size. Our sample including healthy subjects also demonstrated a decrease in regional size with increasing age, as documented in the literature.[27]

Next, we assessed the impact of the patient's demographic characteristics on regional brain volume. Regardless of the patients' clinical presentation, increased time
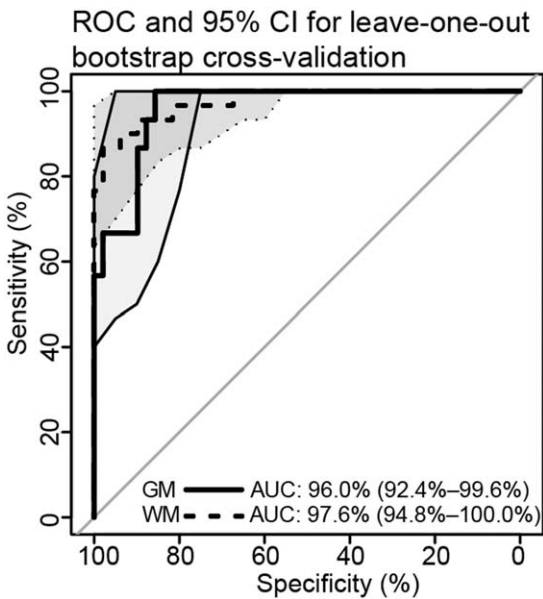
FIGURE 4: ROC curves show the trade-off between sensitivity and specificity for the leave-one-out bootstrap cross-validation for the classification of UWS and MCS patients based on grey and white matter regional volumes. GM = gray matter; WM = white matter.
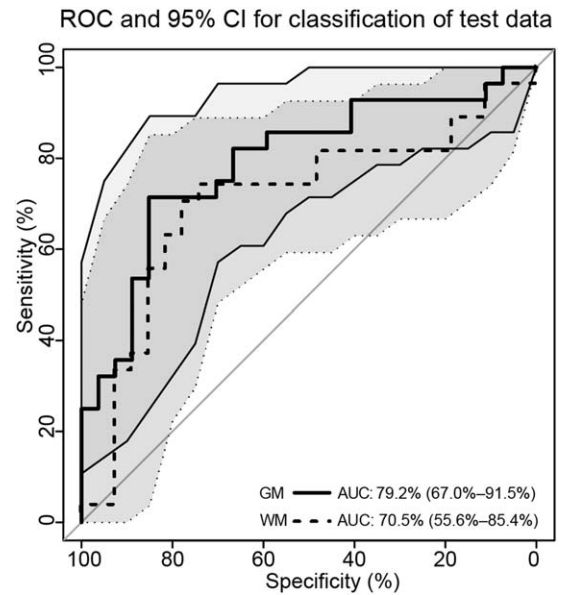
FIGURE 6: ROC curves present the trade-off between sensitivity and specificity for the classification of UWS and MCS patients from the independent test data sets (ie, Salzburg and Paris) using grey and white matter regional volumes. GM = grey matter, WM = white matter.

since brain injury negatively affected regional GM and WM volume. Nonetheless, unconscious patients show more severe atrophy over time when compared to

patients with some level of residual consciousness (in GM for MCS, in WM and GM for conscious patients). These results highlight the fact that patients sustaining
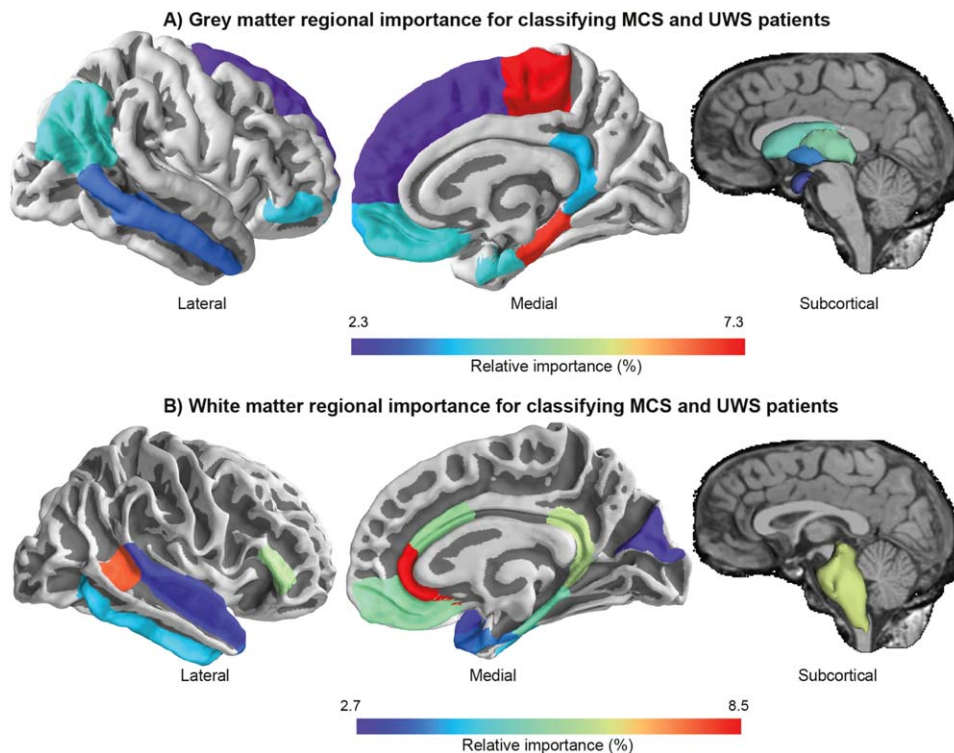

FIGURE 5: Gray (A) and white (B) matter regional importance for the classification of UWS and MCS patients. This parameter identifies the influence of a single feature on the classification result as compared to the other features as presented by the weight of the stump. Only regions with classification importance higher than chance (2.33% for gray matter and 2.66% for white matter) are displayed in color according to their importance. MCS = minimally conscious state, UWS = unresponsive wakefulness syndrome.

awareness are more likely to preserve a larger brain volume over time, whereas unconscious patients are prone to substantial GM and WM loss. Our results in brain-injured patients are in agreement with those of Rubeaux and colleagues,[28] who observed a slower atrophy of the thalamus in MCS patients than in UWS patients. Maintenance of brain volume is also associated with a better prognosis in the acute phase of brain injury.[29] This suggests a "use it or lose it" principle in which activated neurons seem to better withstand neurodegeneration and atrophy, as observed in Alzheimer's disease as well.[30] Two famous postmortem case studies in persistent UWS patients support our finding of more severe atrophy with increasing time postinjury. Karen Ann Quinlan's brain weighed 66% (835g) of the normal 1.3kg after 10 years of UWS,[31] whereas Terri Schiavo's brain showed more atrophy and weighed only 47% (615g) after 15 years of UWS. Similar autopsy studies of MCS patients to confirm our findings are lacking, however.

We did not observe a difference in mean regional volume between UWS and MCS patients, though previous findings suggest that regional differences exist.[12] Indeed, we identified regional relevance to discriminate conscious and unconscious brains with above-chance-level accuracy. This suggests that specific brain areas', rather than the brain in general, structural integrity is important for consciousness. A congruence between the T1-based classification and the CRS-R diagnosis of at least 87% was found, emphasizing the validity of our results. This exceeds previously published PET-based classification accuracies, that is, 83% using cortical variance in glucose uptake,[10] 85% obtained with a similar processing approach as adopted in this article,[9] as well as 87% using metabolism of the best preserved hemisphere.[32] When the T1- and PET-based diagnosis are combined, we find a false-negative rate of 2% and 4% for GM and WM, respectively, suggesting that combining different modalities is a powerful tool to detect residual consciousness. When considering both techniques, the false-positive rate increases, especially in UWS patients. However, these patients (also referred to as MCS* patients) might not show behavioral signs of consciousness, yet results of paraclinical assessments (such as neuroimaging and neurophysiology) are indicative for the presence of residual consciousness.[33] In clinical practice it is important to identify these patients because the diagnosis has significant implications for clinical management, treatment, and prognosis.[34,35]

Good classification accuracy was observed for the test data, with an AUROC of 79% (95% CI [67-91.5]) for GM and 70% (95% CI [55.6-85.4]) for WM. These performances were achieved despite differences in etiology and time post-injury (which are known to have divergent effects on structural integrity[12,36]), number of behavioral assessments (which could influence the final diagnosis of DOC patients[4]), and MRI scanners (which might influence the preprocessing, even if Freesurfer is relatively stable across different MRI scanners[37,38]). This demonstrates the potential of the proposed method in different clinical settings.

From our analysis, it appears that the most informative GM regions include the (para-) hippocampal area and entorhinal cortex. These regions are important for the formation of episodic memory, permit rapid processing enabling contextualization of events in the world, and are well located to bind associations from other cortical streams, suggesting a key function for consciousness.[39] We have found these regions to be severely damaged in persistent UWS patients and to a lesser extend in MCS patients, which is in line with observations in postmortem studies.[40] Other areas that seem to play a key role in the classification protocol are the thalamus and caudate, which are associated to recovery of motor and cognitive function in the mesocircuit hypothesis.[41] The thalamus receives sensory information and redirects the excitatory input to the (frontal) cortex, which, in turn, drives neurons in the caudate (part of the striatum). Generally, the caudate indirectly excites the thalamus, but without sufficient excitation in the striatum the thalamus is inhibited, which may lead to the underlying malfunction of this circuit in DOC patients.[41] Atrophy in the thalamus has been correlated to CRS-R total scores, but could not be related to clinical status directly.[11] Our findings provide direct evidence for a possible relationship between brain structure and pathological states of consciousness. Last, frontal regions together with inferior parietal regions and cingulate areas bear high classification relevance. These regions are all part of the default mode network (DMN), a complex of regions believed to have a prominent role in impaired consciousness. Especially these regions show severe abnormalities in glucose metabolism,[42] structural WM,[43] and functional connectivity,[44,45] that is, linearly related to consciousness levels in DOC patients. The presented results suggest that these functional changes may have a structural basis in GM volume as well.

WM regions with high discriminatory power include the cingulate area, the (superior) temporal WM, and the brainstem. The temporal cortex is involved in auditory processing. Our results are in line with a recent study indicating that functional connectivity within the auditory network is the most sensitive compared to other functional networks to objectively discriminate between UWS and MCS patients.[46] Mostly, WM regions

important for long-range connectivity are affected in UWS patients as compared to MCS patients, corroborating with the notion that loss of consciousness results from regions functioning in isolation.[47] The UWS can be explained as preserved brainstem function with loss of cortical function,[48] even if islands of preserved function might exist during unconsciousness.[49] Yet, we do observe differences in brainstem volume for UWS and MCS patients. This might be attributed to widespread Wallerian-type degeneration affecting the corticospinal tracts and hence brainstem volume,[50] which might be more prominent in UWS patients than in MCS patients who show a broader spectrum of behavior.

Finally, we would like to mention three possible limitations of this work. First of all, brain atrophy was not assessed in a longitudinal fashion. This would be preferable in order to study the dynamics of volumetric changes on the subject level, and would permit to assess the influence of the clinical evolution on the misclassification rate. Yet, the size of the analyzed sample and the large number of observations within each subject support the validity of our findings. Second, even if our approach does not involve brain function, brain activity depends to some extent on structural integrity,[43] and indeed our results indicate a congruence with glucose uptake. Therefore, we believe our results are clinically relevant. Last, given the nature of the fully automated analysis, we have only included patients with a relatively preserved brain and excluded around one third of the patient data from all three centers. This could bias our results, and appear as a limitation in clinical practice. However, this bias exists for all neuroimaging methods. Moreover, we believe that the current study emphasizes the importance of structural investigation in addition to functional neuroimaging, of which the latter could be difficult to implement in clinical practice given that functional data is heavily confounded by sedation and motion.

To summarize, we evaluated regional GM and WM differences in healthy subjects and patients, showing that brain-injured patients have decreased regional volumes compared to healthy subjects. Within the patient sample, we tested for group differences and have found diminished regional volume in UWS compared to conscious patients. Moreover, there was evidence for a negative relationship between regional brain volume and disease duration regardless of diagnosis, whereas atrophy appears faster in UWS patients than in the MCS (GM) and conscious patients (GM and WM). Finally, regional volume seems to discriminate UWS and MCS patients as well (internal validation GM-AUROC 96% (95% CI [92.4–99.6]), WM-AUROC 97.6% (95% CI [94.8–100]); independent test data GM-AUROC 79.2% (95% CI [67.0–91.5]), WM-AUROC 70% (95% CI [55.6–85.4])). The most important GM regions for classification belong to the DMN and subcortical areas, and relevant WM areas are important for long-range connectivity. Our findings highlight the importance of regional brain volume assessments for clinical evaluation of DOC patients. Therefore, structural MRI quantification should be considered alongside functional neuroimaging to improve the clinical diagnosis of DOC patients.

Author Contributions

J.A., G.F., and S.L. were responsible for study concept and design. J.A., G.F., J.S.C., L.H., C.M., A.D., C.dP., and H.C. contributed to data acquisition and analysis. J.A., G.F., J.S.C., C.dP., L.N., and S.L. contributed to the drafting of the manuscript and/or figures.

The Coma Science Group contributors were responsible for data acquisition and revision of the manuscript.

Coma Science Group collaborators:

Georgios Antonopoulous, MSc,[1] Charlène Aubinet, MSc,[1,2] Geraldine Martens, MSc,[1,2] Audrey Vanhaudenhuyse, Ph.D.,[1,2] Damien Galanaud, M.D.,[8] Benjamin Rohaut, M.D., Ph.D.,[6,9] Eugen Trinka, M.D., Ph.D.,[10] Martin Kronbichler, Ph.D.,[11] Manuel Schabus, Ph.D.,[11] Jean-Flory Luaba Thsibanda, M.D.,[2] Roland Hustinx, M.D., Ph.D.,[2] Claire Bernard, MSc,[2] Murielle Kirsch, M.D.,[2] and Jean-François Brichant, M.D., Ph.D.[2] [8]AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Neuroradiology, Paris, France; [9]Department of Neurology, Columbia University Medical Center, New York, NY; [10]Department of Neurology, Christian Doppler Klinik, Paracelsus Medical University, Salzburg, Austria; and [11]Centre for Cognitive Neuroscience & Department of Psychology, University of Salzburg, Salzburg, Austria

Potential Conflicts of Interest

Nothing to report.

References

1. Giacino JT, Schnakers C, Rodriguez-Moreno D, et al. Behavioral assessment in patients with disorders of consciousness: gold standard or fool's gold? Prog Brain Res 2009;177:33–48.

2. Giacino JT, Kalmar K. Coma Recovery Scale Revised Administration and Scoring Guidelines. Edison, NJ: Center for Head Injuries; 2004.

3. Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: Measurement characteristics and diagnostic utility. Arch Phys Med Rehabil 2004;85:2020–2029.

4. Wannez S, Heine L, Thonnard M, et al. The repetition of behavioral assessments in diagnosis of disorders of consciousness. 2017;81:883–889.

5. Laureys S, Celesia GG, Cohadon F, et al. Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. BMC Med 2010;8:1–4.

6. Giacino JT, Ashwal S, Childs N, et al. The minimally conscious state: Definition and diagnostic criteria. Neurology 2002;58:349–353.

7. Patterson JR, Grabois M. Locked-in syndrome: a review of 139 cases. Stroke 1986;17:758–764.

8. Schnakers C, Vanhaudenhuyse A, Giacino J, et al. Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. BMC Neurol 2009;9:35.

9. Stender J, Gosseries O, Bruno M-A, et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. Lancet 2014;6736:8–16.

10. Stender J, Kupers R, Rodell A, et al. Quantitative rates of brain glucose metabolism distinguish minimally conscious from vegetative state patients. J Cereb Blood Flow Metab 2014;2014:1–8.

11. Lutkenhoff ES, Chiang J, Tshibanda L, et al. Thalamic and extra-thalamic mechanisms of consciousness after severe brain injury. Ann Neurol 2015;78:68–76.

12. Guldenmund P, Soddu A, Baquero K, et al. Structural brain injury in patients with disorders of consciousness: a voxel-based morphometry study. Brain Inj 2016;9052:1–10.

13. Andersen SM, Rapcsak SZ, Beeson PM. Cost function masking during normalization of brains with focal lesions: still a necessity? Neuroimage 2010;53:78–84.

14. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. J Neurotrauma 1998;15:573–585.

15. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. Int J Surg 2014;12:1495–1499.

16. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17:87–97.

17. Fischl B, Salat DH, Van Der Kouwe AJ, et al. Sequence-independent segmentation of magnetic resonance images. Neuroimage 2004;23:69–84.

18. Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans Med Imaging 2001;20:70–80.

19. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA 2000;97:11050–11055.

20. Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution inter-subject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 1999;8:272–284.

21. Hansen TI, Brezova V, Eikenes L, et al. How does the accuracy of intracranial volume measurements affect normalized brain volumes? Sample size estimates based on 966 subjects from the HUNT MRI Cohort. AJNR Am J Neuroradiol 2015;36:1450–1456.

22. Buckner RL, Head D, Parker J, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 2004;23:724–38.

23. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31:968–980.

24. R Core team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2012. Available at: http://www.r-project.org/. 2012. Accessed 02/04/2018.

25. Breiman L. Arcing classifiers. Ann Stat 1998;26:801–849.

26. Fu WJ, Carroll RJ, Wang S. Estimating misclassification error with small samples via bootstrap cross-validation. Bioinformatics 2005;21:1979–1986.

27. Gur RC, Mozley PD, Resnick SM, et al. Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. Proc Natl Acad Sci U S A 1991;88:2845–2849.

28. Rubeaux M, Jayashri J, Gomez F, et al. Thalamic volume as a biomarker for Disorders Of Consciousness. In: SPIE 9287, 10th International Symposium on Medical Information Processing and Analysis. 2015. p. 1–6.

29. Ross DE, Ochs AL, Seabaugh JM, et al. Progressive brain atrophy in patients with chronic neuropsychiatric symptoms after mild traumatic brain injury: a preliminary study. Brain Inj 2012;26:1500–1509.

30. Swaab D. Brain aging and Alzheimer's disease, "wear and tear" versus "use it or lose it." Neurobiol Aging 1991;12:317–324.

31. Kinney H., Korein K, Panigraphy A, et al. Neuropathological findings in the brain of Karen Ann Quinlan. N Engl J Med 1994;330:1469–1475.

32. Stender J, Mortensen KN, Thibaut A, et al. The minimal energetic requirement of sustained awareness after brain injury. Curr Biol 2016;26:1494–1499.

33. Di Perri C, Thibaut A, Heine L, et al. Towards new methods of diagnosis in disorders of consciousness—authors' reply. Lancet Neurol 2016;15:1115–1116.

34. Demertzi A, Racine E, Bruno MA, et al. Pain perception in disorders of consciousness: Neuroscience, clinical care, and ethics in dialogue. Neuroethics 2013;6:37–50.

35. Thibaut A, Bruno MA, Ledoux D, et al. tDCS in patients with disorders of consciousness. Neurology 2014;82:1–7.

36. Van der Eerden AW, Khalilzadeh O, Perlbarg V, et al. White matter changes in comatose survivors of anoxic ischemic encephalopathy and traumatic brain injury: comparative diffusion-tensor imaging study. Neuroradiology 2014;270:506–516.

37. Han X, Jovicich J, Salat D, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 2006;32:180–194.

38. Reuter M, Schumansky NJ, Rosas DH, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 2012;61:1402–1418.

39. Behrendt R. Hippocampus and consciousness. Rev Neurosci 2013;24:239–266.

40. Jellinger K a. Neuropathology of prolonged unresponsive wakefulness syndrome after blunt head injury: review of 100 postmortem cases. Brain Inj 2013;27:917–923.

41. Schiff ND. Recovery of consciousness after brain injury: a mesocircuit hypothesis. 2010;33:1–9.

42. Thibaut A, Bruno M-A, Chatelle C, et al. Metabolic activity in external and internal awareness networks in severely brain-damaged patients. J Rehabil Med 2012;44:481–516.

43. Annen J, Heine L, Ziegler E, et al. Function—structure connectivity in patients with severe brain injury as measured by MRI-DWI and FDG-PET. Hum Brain Mapp 2016;37:3707–3720.

44. Demertzi A, Gómez F, Crone JS, et al. Multiple fMRI system-level baseline connectivity is disrupted in patients with consciousness alterations. Cortex 2014;52:35–46.

45. Vanhaudenhuyse A, Noirhomme Q, Tshibanda LJ, et al. Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. Brain 2010;133:405–422.

46. Demertzi A, Antonopoulos G, Heine L, et al. Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. Brain 2015;138:2619–31.

47. Rosanova M, Gosseries O, Casarotto S, et al. Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. Brain 2012;135:1–13.

48. Jennett B. A quarter century of the vegetative state: an international perspective. J Head Trauma Rehabil 1997;12:1–12.

49. Schiff ND, Ribary U, Moreno DR, et al. Residual cerebral activity and behavioural fragments can remain in the persistently vegetative brain. Brain 2002;125:1210–1234.

50. 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.