



# The power of sample size through a multi-scanner approach in MR neuroimaging regression analysis: evidence from Alzheimer's disease with and without depression

Efstratios Karavasilis<sup>1,2</sup> · Theodore P. Parthimos<sup>3</sup> · John D. Papatriantafyllou<sup>3</sup> · Foteini Christidi<sup>4</sup> · Sokratis G. Papageorgiou<sup>5</sup> · George Kapsas<sup>6</sup> · Andrew C. Papanicolaou<sup>7,8,9</sup> · Ioannis Seimenis<sup>2</sup>

Received: 23 May 2018 / Accepted: 27 April 2019 / Published online: 3 May 2019  
© Australasian College of Physical Scientists and Engineers in Medicine 2019

## Abstract

The inconsistency of volumetric results often seen in MR neuroimaging studies can be partially attributed to small sample sizes and variable data analysis approaches. Increased sample size through multi-scanner studies can tackle the former, but combining data across different scanner platforms and field-strengths may introduce a variability factor capable of masking subtle statistical differences. To investigate the sample size effect on regression analysis between depressive symptoms and grey matter volume (GMV) loss in Alzheimer's disease (AD), a retrospective multi-scanner investigation was conducted. A cohort of 172 AD patients, with or without comorbid depressive symptoms, was studied. Patients were scanned with different imaging protocols in four different MRI scanners operating at either 1.5 T or 3.0 T. Acquired data were uniformly analyzed using the computational anatomy toolbox (CAT12) of the statistical parametric mapping (SPM12) software. Single- and multi-scanner regression analyses were applied to identify the anatomical pattern of correlation between GM loss and depression severity. A common anatomical pattern of correlation between GMV loss and increased depression severity, mostly involving sensorimotor areas, was identified in all patient subgroups imaged in different scanners. Analysis of the pooled multi-scanner data confirmed the above finding employing a more conservative statistical criterion. In the retrospective multi-scanner setting, a significant correlation was also exhibited for temporal and frontal areas. Increasing the sample size by retrospectively pooling multi-scanner data, irrespective of the acquisition platform and parameters employed, can facilitate the identification of anatomical areas with a strong correlation between GMV changes and depression symptoms in AD patients.

**Keywords** Sample size · Neuroimaging · Regression · Alzheimer's disease · Depression

✉ Efstratios Karavasilis  
stratoskaravasilis@yahoo.gr

<sup>1</sup> Second Department of Radiology, University General Hospital 'Attikon', Medical School, National and Kapodistrian University of Athens, 19 Papadiamantopoulou Street, Athens, Greece

<sup>2</sup> Department of Medical Physics, Medical School, Democritus University of Thrace, Alexandroupolis, Greece

<sup>3</sup> 3rd Age Day Care Center IASIS, Glyfada, Greece

<sup>4</sup> First Department of Neurology, Aeginition Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece

<sup>5</sup> Second Department of Neurology, University General Hospital 'Attikon', Medical School, National and Kapodistrian University of Athens, Athens, Greece

<sup>6</sup> Department of Radiology, 424 General Military Hospital, Thessaloniki, Greece

<sup>7</sup> Division of Clinical Neurosciences, Department of Pediatrics, University of Tennessee Health Science Center, Memphis, USA

<sup>8</sup> Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, USA

<sup>9</sup> Neuroscience Institute, Le Bonheur Children's Hospital, Memphis, USA

## Introduction

Advanced magnetic resonance (MR) neuroimaging techniques are increasingly used to aid clinicians in estimating the effect of aging in healthy control groups [1–4], as well as in the early or differential diagnosis of several neurodevelopmental or neurodegenerative disorders, including schizophrenia, epilepsy, multiple sclerosis, and Alzheimers' disease (AD) [5–10]. One of the most prominent contributions of MR neuroimaging is volumetric assessment, often used for the detection of brain volume alterations in clinical groups and in the identification of correlations between volumetric changes and clinical indices. The employed analysis technique is called voxel-based morphometry (VBM) and is based on the statistical parametric mapping of three-dimensionally (3D) acquired, high resolution, T1-weighted MR images. The inconsistency of results often reported in the literature partially originates from the use of small sample sizes [11]. This limitation cannot be always overcome by the optimization of the acquisition protocol and the fine-tuning of the post-processing analysis and, thus, the multi-scanner approach offers an attractive alternative [11].

Prospective, multi-scanner studies, however, necessitate a well-conceived and executed research design, hindering their practical implementation. Nevertheless, 3D, high resolution, T1-weighted images are routinely used, regardless of the specific scanner and imaging protocol employed, for global or focal gray matter volumetry in patients with neurodevelopmental or neurodegenerative disorders. This point, along with the fact that quantitative neuropsychological assessment is usually performed in these patients, suggest that the creation of an appropriate retrospective “multi-scanner” dataset can constitute a realistic alternative in the investigation of subtle VBM differences and their correlation to neuropsychological indices.

With the advantage of retrospectively pooling multi-scanner data to expand the sample size come important drawbacks. It is widely accepted that experimental variance in quantitative MR imaging studies increases when data are acquired by multiple scanners and protocols [11, 12]. Different scanner platforms and static magnetic field strengths, field inhomogeneities, type of receiving coils, variation in patient positioning and geometry distortions due to non-linear gradients are some of the factors which have been considered as sources of variance in multi-scanner studies [13–15]. A phantom-based quality control procedure, as well as the scanning of a small group of volunteers in all available MR scanners, have been suggested [15] as methodological strategies to calibrate multi-scanner studies. However, conducting such strategies is a complex and time-consuming undertaking. On

the other hand, MR scanner related inherent variations are averaged out in the multi-scanner setting, whilst the effect of systematic or random errors associated with a single scanner is minimized.

The current study focuses on the potential correlations between grey matter volume (GMV) and depression symptoms in AD patients. To investigate how the sample size affects the outcome of regression analysis between gray matter loss and severity of comorbid depression, for which the neuroanatomical substrate is still debatable [16], we retrospectively studied a cohort of 151 AD patients with and without depression. Patients were scanned with different imaging protocols in four different MRI scanners operating at either 1.5 T or 3.0 T. Whole brain regression analysis of single- and multi-scanner datasets was employed to identify anatomical areas of correlation between severity of depression symptomatology and grey matter volume loss.

## Methods

### Participants

One hundred seventy two (172) AD patients with and without comorbid depression were initially registered in this retrospective study. The patients were recruited from an outpatient Memory clinic and a Day Care Center for 3rd Age. The inclusion criteria were (a) clinical diagnosis of AD according to established National Institute on Aging/Alzheimer's Association (NIAA) criteria [17] (b) Greek as native language and (c) Mini Mental State Examination (MMSE) score  $\geq 17$  (since individuals with an MMSE score lower than 15 may be regarded as non-reliable in self-assessment tests [18–20]). Participants were excluded if they had a premorbid central nervous system disorder, if they had a premorbid major depression or other significant psychiatric disorder as defined by the Diagnostic and Statistical Manual of Mental Disorders [21] and if they had an active or current alcohol, drug or other controlled substance abuse history.

Informed consent to participate in the study was obtained from all participants. The study was conducted in compliance with the regulations of the local ethics committee and in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

### Neurocognitive and neuropsychiatric assessment

Depression was evaluated using the Geriatric Depression Scale (GDS) [22] for participants older than 65 years old and the Zung Depression Scale [23] for participants up to the age of 65 years ( $n = 19$ ). GDS was also administered to patients younger than 65 years old to facilitate correlation with neuroimaging data. A GDS score of  $\leq 10$  is considered as the

cut-off point for depression (mild depression: 10–19; severe depression: 20–30). With regard to the Zung Depression Scale, a score equal to or higher than 50 indicates depression (mild: 50–59, moderate: 60–69, severe  $\geq 70$ ). Cognitive assessment included the Addenbrooke's Cognitive Examination Revised (ACE-R) [24], the Frontal Assessment Battery [25], the Clock Drawing Test [24] and verbal fluency (animals and words/min). For the purpose of the current neuroimaging analysis, only the ACE-R and GDS results were considered.

### MRI data acquisition protocol

All participants underwent a brain MRI exam as part of their routine clinical evaluation within one month from the time of neurocognitive and neuropsychiatric assessment. In all cases, the brain MRI protocol included a 3D, high spatial resolution T1-weighted (3D HR-T1) gradient echo pulse sequence for the acquisition of detailed anatomical images. MRI scans were performed in four different MR scanners. Therefore, acquisition parameters varied depending on the MR scanner and exact protocol used. Nevertheless, all employed acquisition protocols met the minimum requirements for obtaining images suitable for VBM analysis (reconstructed voxel size  $\leq 1.2 \text{ mm} \times 1.2 \text{ mm} \times 1.2 \text{ mm}$ ). The main acquisition/reconstruction parameters and the corresponding static magnetic field strengths are given in Table 1. All acquired images were screened by an experienced neuro-radiologist using standard neuroradiological criteria for the detection of pathologies or abnormalities, including white matter hyperintensities, and the presence of image artifacts (e.g., due to gross motion).

### MRI data preprocessing

Volumetric analysis was performed using the computational anatomy toolbox (CAT12), a toolbox of statistical parametric mapping (SPM12; Wellcome Department of Cognitive Neurology, <https://www.fil.ion.ucl.ac.uk/spm/software/spm12>) implemented on MATLAB R2015b (The MathWorks, Natick, USA). All 3D-HR T1 images were initially segmented into GM, white matter (WM) and cerebrospinal

fluid, and then were normalized using Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL). For registration purposes, six iterations and an existing DARTEL template in MNI space, derived from 555 healthy controls of the IXI-database (<https://www.brain-development.org>), were employed. During this registration procedure, local GM and WM volumes are conserved by modulating the image intensity of each voxel by the Jacobian determinants of the computed deformation fields. Registered image and preprocessing parameters exported were quantitatively assessed and data with weighted overall quality measure (IQR) lower than C+ were excluded from further analysis. The remaining normalized and modulated GM images were smoothed with an 8 mm full-width-at-half-maximum isotropic Gaussian kernel via a standard module of SPM.

Anatomical region volumes were calculated using script `get_totals.m` (developed by Ged Ridgway), VBM12 and WFU\_Pickatlas MATLAB toolboxes, according to the following steps: (1) Masks of the anatomical region of interest were created using WFU\_Pickatlas toolbox (<https://www.Fmri.wfubmc.edu>); (2) the masks' voxel size was interpolated to the voxel size of the modulated normalized segmented data using `co-register (re-slice)` module of SPM12; (3) volumes of anatomical regions of interest were calculated using script `get_totals.m`.

### Statistical analysis

Pre-processed images were fed into SPM12 statistical models. Whole-brain correlation analysis between GM volume and GDS was performed using the "multiple regression" design function of SPM12. To this purpose, GDS was used as covariate of interest, whilst age, gender, years of education, total intracranial volume (TIV), age onset and ACE-R were used as confounding variables to account for any potentially contributing effect on the pattern of focal GM changes [26–28]. The above statistical design was performed for (1) participants scanned in each MRI scanner separately, (2) all study participants independent of MRI scanner, (3) participants scanned at 1.5 T scanners only and (4) participants scanned at 3.0 T scanners only. In cases (2), (3) and (4), the MRI scanner was used as a control covariate.

**Table 1** Acquisition parameters and field strength of scanners employed

# of MR scanner	TR/TE(ms)	Flip angle (°)	Acquisition matrix	Acquired/reconstructed slice thickness (mm)	Static magnetic field strength (Tesla)
1	9.3/3.7	15	320×320	1.4/0.7	1.5
2	18/4.9	25	256×200	1.0/1.0	3.0
3	7.1/3.1	15	256×224	1.4/0.7	1.5
4	9.9/3.7	7	256×237	1.0/1.0	3.0

MR magnetic resonance, TR repetition time, TE echo time, ms milliseconds, mm millimeters

The statistical threshold was set at  $p < 0.05$ , applying family-wise error (FWE) correction for multiple comparisons. A more liberal criterion was also applied ( $p < 0.001$  uncorrected, extended threshold: expected voxels per cluster). Anatomical regions of interest covering the entire volumes of clusters were defined using the WFU PickAtlas tool of SPM [29, 30] and Automated Anatomical Labeling (AAL) [31].

## Results

### Demographic and clinical data of participants

Twenty one MR image datasets of both depressed and non-depressed AD patients were excluded from the analyzed sample due to the presence of abnormalities and/or severe artifacts (qualitative assessment by the neuroradiologist) or because of inadequate image quality (IQR < C+, quantitative assessment by the CAT12 toolbox). Table 2 presents demographic data and clinical characteristics for the four groups scanned in the four different MR scanners.

### Regression analysis

A common anatomical pattern of correlation between GM volume loss and increased GDS score, mostly involving sensorimotor areas, was identified in all patient subgroups imaged at different scanners (Fig. 1a). Additionally, whole brain regression analysis revealed significant correlations in several frontal, temporal, cerebellar and basal ganglia structures in some but not all scanner subgroups (Table 2).

When combining imaging data from all scanners, the detected correlation pattern of GM volume loss and comorbid depression severity confirmed the above findings and appeared to extend towards the temporal and frontal lobes (Fig. 1b). Furthermore, enlargement of the sample size

clearly promotes statistical significance, since thalamic and sensorimotor clusters survived FWE correction at  $p < 0.05$  in the combined analysis of the total study cohort (Table 3).

Figure 2 highlights the correlation between GM volume loss and increased GDS score for the patient subgroups scanned at 1.5 T (a) and 3.0 T (b), respectively. A common anatomical pattern of strong correlation, mainly involving sensorimotor areas and bilateral thalamus, was revealed in both subgroups. However, regression analysis of the 1.5 T dataset also exhibited significant correlations in frontal, temporal and cerebellar areas, as well as in the right insula, while a significant correlation for the 3.0 T dataset was shown in occipital areas (Table 3).

Table 4 presents the anatomical regions, per scanner, for which GMV loss was associated with increased GDS. Results from various scanners are compatible, although between-scanner variations, also resulting from unmatched patient cohorts, are noticeable.

## Discussion

The main goal of this study was to increase the sample size by retrospectively pooling multi-scanner neuroimaging data, regardless of acquisition field strength and imaging parameters, to facilitate the identification of brain areas with a significant correlation between depressive symptoms severity and brain atrophy in AD. To this purpose, whole brain regression analysis models were applied without a priori theoretical preconceptions as to what brain regions might be affected. Multi-scanner outcomes were compared against single-scanner and single-field findings. Current results suggest that the practical approach of employing large-sample, multi-scanner analysis is more robust to identify brain areas with correlation between GM loss and depression severity, compared to the analysis of a relatively more homogeneous and optimized, yet smaller, dataset acquired from a single

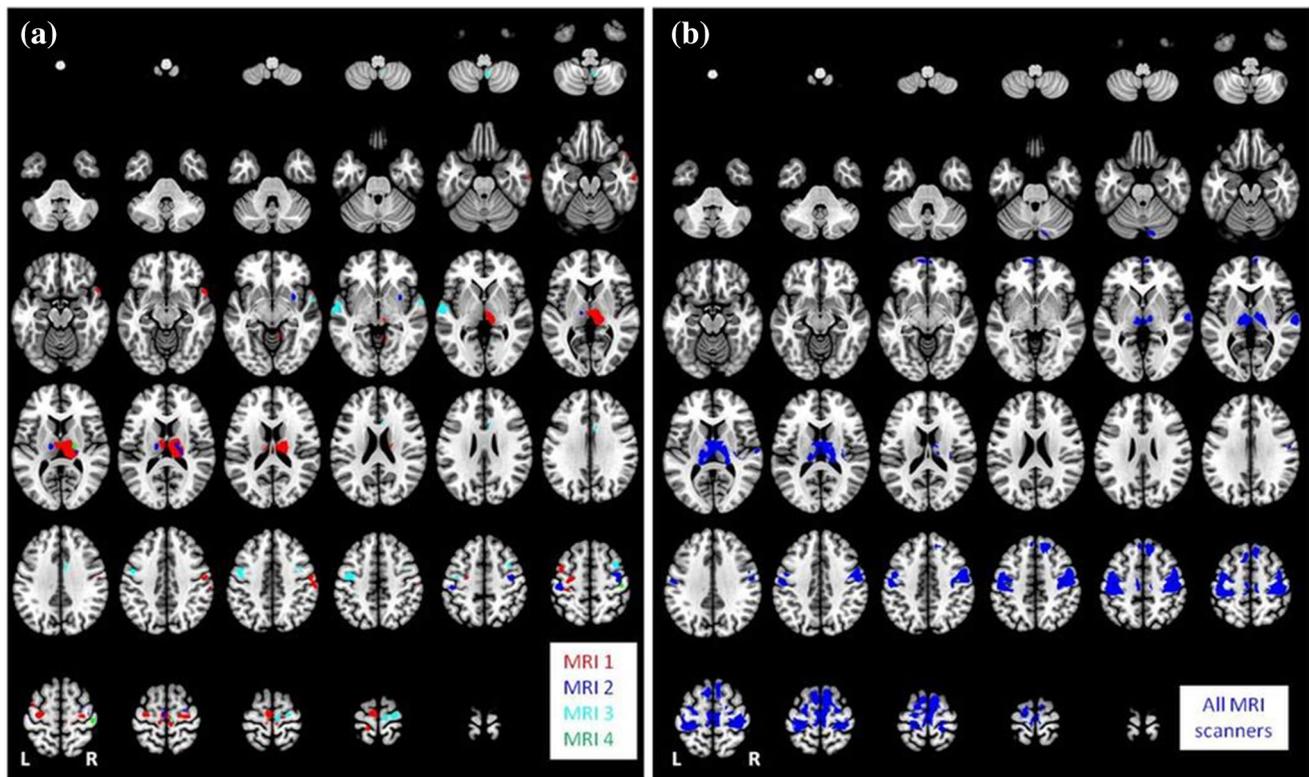
**Table 2** Demographic data and clinical characteristics of patient groups analyzed

Variables	MRI scanner				Statistics
	1 (n=38)	2 (n=61)	3 (n=28)	4 (n=24)	
Age (y)	71.7 ± 9.1	74.2 ± 8.2	74.7 ± 6.7	76.1 ± 6.5	ns
Gender (M/F)	8/30	20/41	10/18	3/21	ns
Education (y)	11.7 ± 3.8	10.2 ± 4.3	11.5 ± 4.1	8.6 ± 5.2	F = 3.135; p = 0.027 <sup>a</sup>
ACE-R	67.0 ± 12.7	62.8 ± 13.1	69.4 ± 10.6	58.9 ± 9.4	F = 4.205; p = 0.007 <sup>b</sup>
Age onset	68.2 ± 9.8	71.8 ± 8.6	72.0 ± 6.8	73.2 ± 7.0	ns
GDS	6.3 ± 5.4	7.7 ± 6.0	6.4 ± 5.5	7.4 ± 6.1	ns
Depression (N/Y)	30/8	41/20	19/9	18/6	ns

n number of patients, y years, M/F male/female, ACE-R Addenbrooke's Cognitive Examination Revised, GDS Geriatric Depression Scale, N/Y no/yes, ns non-significant at  $p < 0.05$

<sup>a</sup> MRI 1 > MRI 4 ( $p = 0.006$ ); MRI 3 > MRI 4 ( $p = 0.017$ )

<sup>b</sup> MRI 1 > MRI 4 ( $p = 0.011$ ); MRI 3 > MRI 2 ( $p = 0.018$ ); MRI 3 > MRI 4 ( $p = 0.002$ )



**Fig. 1** Association between grey matter volume loss and Geriatric Depression Scale score for **a** individual scanner subgroups ( $p < 0.001$ , uncorrected, extended threshold:  $k = 75$  (MRI 1),  $k = 88$  (MRI 2),

$k = 73$  (MRI 3),  $k = 59$  (MRI 4) voxels) and **b** all participants independent of MRI scanner used ( $p < 0.05$ , FWE-corrected)

MR scanner or a limited number of scanners operating at the same field strength.

In the context of MR neuroimaging research, between-center variations are likely to include the use of different scanner manufacturers, models and releases, coil configurations, field inhomogeneity levels and acquisition or reconstruction parameters inherent to individual scanners [32, 33]. Inevitably, these variations could extenuate subtle statistical differences. Indeed, inter-scanner studies report major inconsistencies in volumetric measurements across different scanners [34, 35]. Discrepancies between results reported in single-scanner studies can be attributed to small sample sizes, specific group clinical characteristics and medications, inherent scanner-related artifacts, as well as to variable preprocessing parameter settings and analysis methods. Additionally, acquisition magnetic field strength may introduce a considerable bias in volumetric measurements [36, 37], also due to RF transmission field inhomogeneity which enhances signal and contrast non-uniformity in the resulting images [38]. Previous studies have detected total GM and WM volume differences [36, 37], as well as regional differences [38, 39]. Regarding total GM and WM volumes, this study didn't reveal any significant differences between patients scanned at 1.5 T

compared to those scanned at 3.0 T. The 1.5 T group, however, exhibited a wider range of anatomical regions with significant correlation, compared to the 3.0 T group, whilst the relevant anatomical pattern better matched that of the total patient cohort.

The well documented inter-scanner variation in volumetric measurements is comparable to that corresponding to within-scanner measurements [15]. Specifically, significant within-scanner volumetric deviations have been found in repetitive measurements of the same subject in the same MR scanner [35]. This compromised reproducibility can be attributed to system-related factors, such as variable shimming and resonance frequency optimization settings, and/or subject-related factors, including non-identical subject positioning that affects magnetic field homogeneity. In this sense, a multi-scanner approach does not necessarily lead to an increase in the observed deviation between measured volumes. On the other hand, current results highlight that the increase in sample size associated with a multi-scanner analysis has a positive outcome in the achieved statistical confidence. As shown in Table 3, a distinct anatomical pattern with correlations, mainly involving sensorimotor and thalamic regions, survived after correction for multiple comparisons.

**Table 3** Anatomical regions for which reduced gray matter volume was associated with increased Geriatric Depression Scale score in all participants, in individual scanner subgroups, as well as at different magnetic field strengths

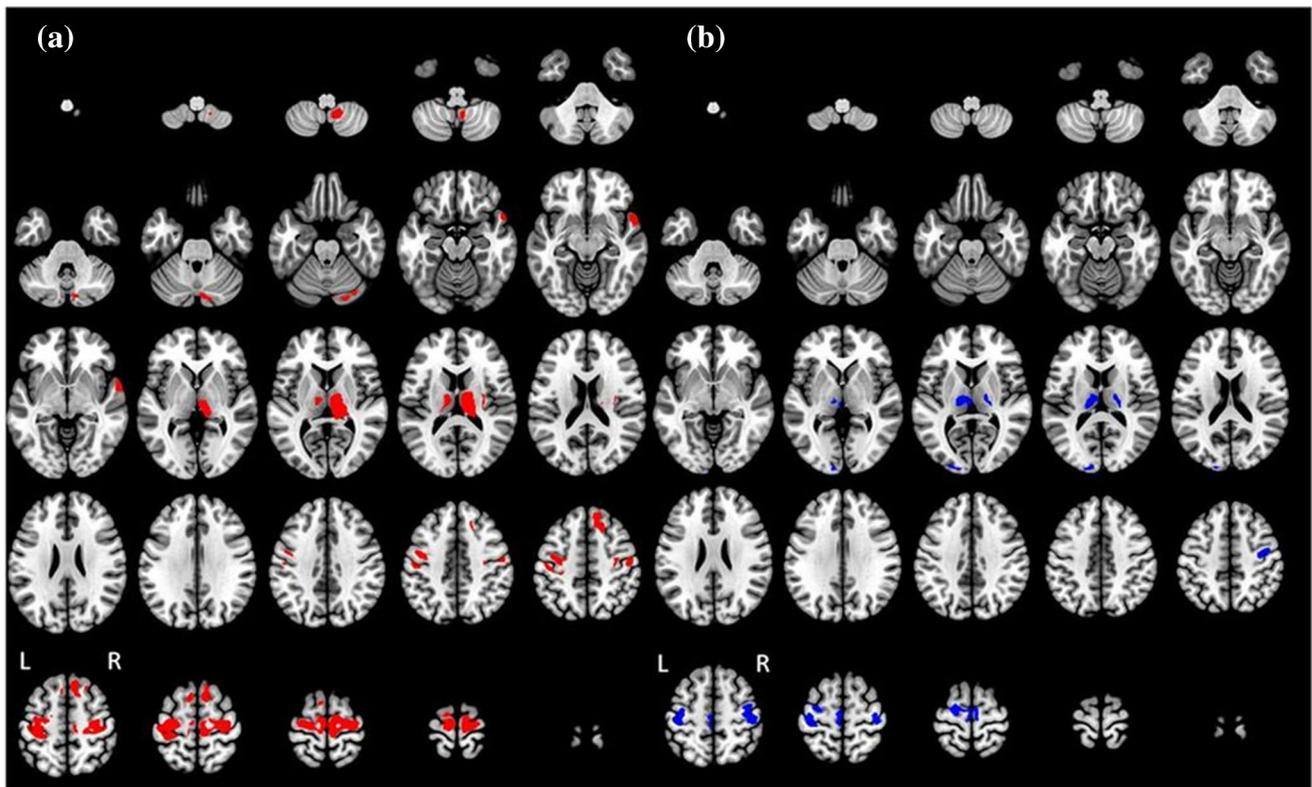
Anatomical labeling	MRI scanners					1.5 T	3.0 T
	All	1	2	3	4		
Thalamus_R	√*	√*	√		√	√*	√
Thalamus_L	√*	√	√			√	√*
Postcentral_R	√*	√			√	√	√
Postcentral_L	√*	√	√	√		√	√
Precentral_R	√*	√	√	√	√	√*	√
Precentral_L	√*	√	√	√		√	√
Supp_Motor_Area_R	√*	√		√		√	√
Supp_Motor_Area_L	√*	√	√			√*	√
Paracentral_Lobule_R	√*	√		√		√	
Paracentral_Lobule_L	√*	√	√			√*	√
Frontal_Sup_Med_R	√						
Frontal_Sup_Med_L	√						
Frontal_Sup_R	√		√			√	
Frontal_Sup_L	√						
Frontal_Mid_R	√		√	√			
Frontal_Med_Orb_R	√						
Frontal_Med_Orb_L	√						
Frontal_Sup_Med_R	√					√	
Temporal_Sup_R	√			√		√	
Temporal_Sup_L				√			
Temporal_Mid_R		√					
Temporal_Mid_L				√			
Temporal_Pole_Sup_R		√		√		√	
Cingulum_Mid_R				√			
Cingulum_Ant_R				√			
Insula_R	√					√	
Putamen_R			√				
Cerebellum_Crus1_R	√					√	
Cerebellum_Crus2_R						√	
Cerebellum_4_5_R		√					
Cerebellum_9_R				√		√	
Occipital Mid L							√
Occipital Sup L							√

Checkmark symbol denotes significance at  $p < 0.001$ , uncorrected (extended threshold:  $k = 75$  (MRI 1),  $k = 88$  (MRI 2),  $k = 73$  (MRI 3),  $k = 59$  (MRI 4),  $k = 90$  (1.5 T),  $k = 96$  (3.0 T) voxels), whilst star symbol denotes significance at  $p < 0.05$ , FWE-corrected

L/R left/right, *Supp\_Motor\_Area* supplementary motor area, *Sup* superior, *Mid* middle, *Med* medial, *Ant* anterior, *Orb* orbital

The robustness of the followed multi-scanner approach is not solely based on the higher level of statistical significance attained, but it is also supported by the identification of additional areas with significant correlations. Multi-scanner data analysis revealed anatomical regions of significant correlation between GM volume loss and depression severity, which were not identified in single-scanner or single-field analyses. The detected areas, mainly involving mediofrontal and orbitofrontal cortex, right insula and right cerebellar crus, have also been identified as regions prone to volume

reduction in late life depression [40, 41]. In addition, an increased sample size, by means of a multi-scanner dataset, seems to minimize the risk of false positive findings resulting from single-scanner related effects in volumetric regression analysis. As evidenced in Table 3, the involvement of temporal and limbic regions mainly detected in a specific single-scanner (#3) analysis was not confirmed by other single-scanner or multi-scanner analyses. Although ACE-R score was treated as a confounding variable, it is postulated that this finding is partly related to the relatively high



**Fig. 2** Association between grey matter volume loss and Geriatric Depression Scale score for patients scanned at **a** 1.5 T magnetic field strength ( $p < 0.001$ , uncorrected, extended threshold  $k = 90$  voxels) and **b** 3.0 T magnetic field strength ( $p < 0.001$ , uncorrected, extended threshold  $k = 96$  voxels)

**Table 4** Volumes (ml) of anatomical regions for which reduced gray matter volume was associated with increased Geriatric Depression Scale score in all participants for each scanner

Anatomical labeling	Volumes (ml) in each MRI scanner			
	#1	#2	#3	#4
Thalamus_R	3.94 ± 0.51	3.66 ± 0.59	3.52 ± 0.26	3.04 ± 0.67
Thalamus_L	3.83 ± 0.57	2.87 ± 0.66	3.43 ± 0.64	3.47 ± 0.48
Postcentral_R	8.21 ± 0.75	7.56 ± 1.19	7.88 ± 0.98	7.81 ± 1.11
Postcentral_L	9.08 ± 0.97	8.11 ± 1.09	8.68 ± 1.32	8.38 ± 0.99
Precentral_R	7.61 ± 0.76	6.57 ± 1.17	7.15 ± 1.14	7.33 ± 1.08
Precentral_L	8.54 ± 1.13	7.43 ± 1.23	8.05 ± 1.43	7.89 ± 1.18
Supp_Motor_Area_R	6.24 ± 0.84	5.66 ± 0.93	6.00 ± 1.16	5.70 ± 0.87
Supp_Motor_Area_L	5.04 ± 0.60	4.62 ± 0.76	4.97 ± 0.90	4.79 ± 0.75
Paracentral_Lobule_R	2.05 ± 0.32	1.81 ± 0.32	1.90 ± 0.36	1.83 ± 0.35
Paracentral_Lobule_L	3.00 ± 0.50	2.43 ± 0.57	3.06 ± 0.61	2.96 ± 0.47
Frontal_Sup_R	9.35 ± 1.33	8.41 ± 1.51	9.13 ± 1.53	8.51 ± 1.20
Frontal_Mid_R	12.78 ± 1.68	12.27 ± 1.96	12.16 ± 1.68	12.09 ± 2.12
Temporal_Sup_R	6.98 ± 0.90	7.24 ± 0.95	7.05 ± 1.05	7.13 ± 1.27

Only the anatomical regions which survived the statistical threshold of  $p = 0.05$  FWE corrected or were detected as statistical significant ( $p < 0.001$  uncorrected) in at least two MRI scanners are shown

L/R left/right, Supp\_Motor\_Area supplementary motor area, Sup superior, Mid middle, Med medial, Ant anterior, Orb orbital

cognitive outcome of this specific patient group (Table 2). In any case, the employment of multi-scanner data averages out different scanner effects, thus attenuating potential single-scanner bias or systematic error [11].

This work provides evidence that multi-institution and multi-scanner morphometric data can be of some interest, even when they are acquired without using an optimized imaging protocol and a standardized procedure as in the Alzheimer's disease neuroimaging initiative (ADNI) study [42, 43]. Although the considerable differences between systems may render pooling absolute measures and ratios problematic [32], current results suggest that correlation indices are comparable across different systems and protocols. It is inevitable, however, that variations in imaging settings or parameters will adversely affect multi-scanner results under certain circumstances. In such a case, one may use a phantom or a small common cohort, if feasible, to unveil specific scanners or institutions that may be responsible for significant data variability. Such an approach could allow identification of inconsistent data, whilst if a suitable phantom is utilized, post-imaging processing could be employed to eliminate geometric distortions.

Several caveats of the current study are noteworthy. Despite the adopted cut-off MMSE value of 17, the fact that the depression tests employed are self-assessment surveys adds significant variance. Although visual and quantitative quality control of imaging data was performed, we were not able to directly validate scanner-related parameters, such as magnetic field homogeneity and gradient linearity. No attempt was made to control for the effect of antidepressant or other medications. Since however medications were too diverse in the studied cohort, a rigorous statistical evaluation was impractical. Another limitation is the lack of a single-scanner study comprising of a large sample of AD patients with and without depression in order to establish a ground truth of the neuroanatomical substrate of depression symptoms in AD. Nevertheless, our results are consistent with the findings of another work [44] studying two independent cohorts, the first one consisting of 41 patients scanned with the same system and the second one consisting of 148 ADNI patients imaged at 58 scanners with a standardized protocol optimized to eliminate between-system inconsistencies. The authors found that even mild and subsyndromal depression is associated with cortical thinning in patients with AD and that depression-related cortical changes mainly included temporal and parietal regions, also identified in this work. The involvement of frontal regions, as well as midbrain and hindbrain structures, demonstrated in the current study may be partly due to the severity of depressive symptoms in the studied cohort. The current analyses merely reflect the greater statistical power of an increased subject cohort. Nevertheless, this is a retrospective study of heterogeneous imaging datasets acquired as part of patients' routine

evaluation. To investigate if significant differences exist among single-scanner groups, either a comparison between scanner groups in a significantly larger population needs to be performed or rescans of the same subjects on different scanners are required.

In conclusion, findings of the current study suggest that increasing the patient cohort, by retrospectively pooling multi-scanner data, seems to constitute a practical approach in MR neuroimaging regression analysis and might be superior to the single-scanner scheme involving a homogeneous but limited dataset.

**Acknowledgements** There is not any external funding source for this work and authors do not declare any potential financial conflicts-of-interest.

**Funding** There is not any funding for this study.

## Compliance with ethical standards

**Conflict of interest** All authors declare that they do not have any conflict of interest.

**Ethical approval** All procedures performed in this study which involved human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

1. Mueller SG, Stables L, Du AT et al (2007) Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4T. *Neurobiol Aging* 28:719–726
2. Jack CR Jr, Wiste HJ, Weigand SD et al (2017) Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement* 13:205–216
3. Szentkuti A, Guderian S, Schiltz K et al (2004) Quantitative MR analyses of the hippocampus: unspecific metabolic changes in aging. *J Neurool* 251:1345–1353
4. Goni W, Abe O, Yamasue H et al (2010) Age-related changes in regional brain volume evaluated by atlas-based method. *Neuroradiology* 52:865–873
5. Agosta F, Galantucci S, Filippi M (2017) Advanced magnetic resonance imaging of neurodegenerative diseases. *Neurol Sci* 38:41–51
6. Wright IC, Rabe-Hesketh S, Woodruff PW, David AS, Murray RM, Bullmore ET (2000) Meta-analysis of regional brain volumes in schizophrenia. *Am J Psychiatr* 157:16–25
7. Henze R, Brunner R, Thiemann U et al (2011) Gray matter alterations in first-admission adolescents with schizophrenia. *J Neuroimaging* 21:241–246
8. Scanlon C, Ronan L, Doherty CP et al (2013) MRI-based brain structure volumes in temporal lobe epilepsy patients and their unaffected siblings: a preliminary study. *J Neuroimaging* 23:64–70

9. Chivers TR, Constantinescu CS, Tench CR (2015) MRI-based measurement of brain stem cross-sectional area in relapsing-remitting multiple sclerosis. *J Neuroimaging* 25:1002–1006
10. Retico A, Bosco P, Cerello P, Fiorina E, Chincarini A, Fantacci ME (2015) Predictive models based on support vector machines: whole-brain versus regional analysis of structural MRI in the Alzheimer's disease. *J Neuroimaging* 25:552–563
11. Stonnington CM, Tan G, Klöppel S et al (2008) Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage* 39:1180–1185
12. Steen RG, Hamer RM, Lieberman JA (2007) Measuring brain volume by MR imaging: impact of measurement precision and natural variation on sample size requirements. *Am J Neuroradiol* 28:1119–1125
13. Shuter B, Yeh IB, Graham S, Au C, Wang SC (2008) Reproducibility of brain tissue volumes in longitudinal studies: effects of changes in signal-to-noise ratio and scanner software. *Neuroimage* 41:371–379
14. Han X, Jovicich J, Salat D et al (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32:180–194
15. Jovicich J, Czanner S, Han X et al (2009) MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage* 46:177–192
16. Boccia M, Acierno M, Piccardi L (2015) Neuroanatomy of Alzheimer's disease and late-life depression: a coordinate-based meta-analysis of MRI studies. *J Alzheimers Dis* 46:963–970
17. McKhann GM, Knopman DS, Chertkow H et al (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7:263–269
18. Burke WJ, Nitcher RL, Roccaforte WH, Wengel SP (1992) A prospective evaluation of the Geriatric Depression Scale in an outpatient geriatric assessment center. *J Am Geriatr Soc* 40:1227–1230
19. Conradsson M, Rosendahl E, Littbrand H, Gustafson Y, Olofsson B, Lovheim H (2013) Usefulness of the Geriatric Depression Scale 15-item version among very old people with and without cognitive impairment. *Aging Ment Health* 17:638–645
20. McGivney SA, Mulvihill M, Taylor B (1994) Validating the GDS depression screen in the nursing home. *J Am Geriatr Soc* 42:490–492
21. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Publishing, Arlington, VA
22. Yesavage JA, Brink TL, Rose TL et al (1983) Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 17:37–49
23. Zung WW (1965) A self-rating depression scale. *Arch Gen Psychiatr* 12:63–70
24. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR (2006) The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatr* 21:1078–1085
25. Dubois B, Slachevsky A, Litvan I, Pillon B (2000) The FAB: a frontal assessment battery at bedside. *Neurology* 55:1621–1626
26. Barnes J, Ridgway GR, Bartlett J et al (2010) Head size, age and gender adjustment in MRI studies: a necessary nuisance? *Neuroimage* 53:1244–1255
27. Kostić VS, Agosta F, Petrović I et al (2010) Regional patterns of brain tissue loss associated with depression in Parkinson disease. *Neurology* 75:857–863
28. Rzezak P, Squarzoni P, Duran FL et al (2015) Relationship between brain age-related reduction in gray matter and educational attainment. *PLoS ONE* 10:e0140945
29. Maldjian JA, Laurienti PJ, Burdette JB, Kraft RA (2003) An automated method for neuroanatomic and cytoarchitectonic atlas based interrogation of fMRI Data Sets. *Neuroimage* 19:1233–1239
30. Maldjian JA, Laurienti PJ, Burdette JH (2004) Precentral gyrus discrepancy in electronic versions of the Talairach Atlas. *Neuroimage* 21:450–455
31. Tzourio-Mazoyer N, Landeau B, Papathanassiou D et al (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15:273–289
32. Kruggel F, Turner J, Muftuler LT (2010) Alzheimer's Disease Neuroimaging Initiative. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *Neuroimage*. 49:2123–2133
33. Shokouhi M, Barnes A, Suckling J et al (2011) Assessment of the impact of the scanner-related factors on brain morphometry analysis with Brainvisa. *BMC Med Imaging*. 11:23
34. Focke NK, Helms G, Kaspar S, Diederich C, Tóth V, Dechent P, Mohr A, Paulus W (2011) Multi-site voxel-based morphometry—not quite there yet. *Neuroimage*. 56:1164–1170
35. Huppertz HJ, Kröll-Seger J, Klöppel S, Ganz RE, Kassubek J (2010) Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. *Neuroimage* 49:2216–2224
36. West J, Blystad I, Engström M, Warntjes JB, Lundberg P (2013) Application of quantitative MRI for brain tissue segmentation at 1.5 T and 3.0 T field strengths. *PLoS ONE* 8:e74795
37. Chu R, Tauhid S, Glanz BI et al (2016) Whole brain volume measured from 1.5 T versus 3 T MRI in healthy subjects and patients with multiple sclerosis. *J Neuroimaging* 26:62–67
38. Tardif CL, Collins DL, Pike GB (2010) Regional impact of field strength on voxel-based morphometry results. *Hum Brain Mapp* 31:943–957
39. Brunton S, Gunasinghe C, Jones N, Kempton M, Westman E, Simmons A (2014) A voxel-wise morphometry comparison of the ADNI 1.5 T and ADNI 3.0 T volumetric protocols. *Alzheimer's Dement* 10(4):823
40. Egger K, Schocke M, Weiss E et al (2008) Pattern of brain atrophy in elderly patients with depression revealed by voxel-based morphometry. *Psychiatr Res* 164:237–244
41. Sachs-Ericsson N, Corsentino E, Moxley J et al (2013) A longitudinal study of differences in late- and early-onset geriatric depression: depressive symptoms and psychosocial, cognitive, and neurological functioning. *Aging Ment Health* 17:1–11
42. Mueller SG, Weiner MW, Thal LJ et al (2005) The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am* 15(4):869–877
43. Jack CR, Bernstein MA, Fox NC et al (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Med* 27:685–691
44. Lebedeva A, Westman E, Lebedev AV, Li X, Winblad B, Simmons A, Wahlund LO, Aarsland D (2014) Alzheimer's Disease Neuroimaging Initiative Structural brain changes associated with depressive symptoms in the elderly with Alzheimer's disease. *J Neurol Neurosurg Psychiatr* 85(8):930–935

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.