NeuroImage 118 (2015) 126–132

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Diurnal fluctuations in brain volume: Statistical analyses of MRI from large populations☆

Kunio Nakamura [a,b,*], Robert A. Brown [a], Sridar Narayanan [a,c], D. Louis Collins [a], Douglas L. Arnold [a,c], The Alzheimer's Disease Neuroimaging Initiative

[a] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Quebec H3A 2B4, Canada
[b] Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Avenue, ND20, Cleveland, OH 44195, USA
[c] NeuroRx Research, 3575 Park Avenue, Suite #5322, Montreal, Quebec H2X 4B3, Canada

ARTICLE INFO

Article history:
Received 16 January 2015
Accepted 26 May 2015
Available online 3 June 2015

Keywords:
MRI
Brain atrophy
Diurnal change
Brain parenchymal fraction

ABSTRACT

We investigated fluctuations in brain volume throughout the day using statistical modeling of magnetic resonance imaging (MRI) from large populations. We applied fully automated image analysis software to measure the brain parenchymal fraction (BPF), defined as the ratio of the brain parenchymal volume and intracranial volume, thus accounting for variations in head size. The MRI data came from serial scans of multiple sclerosis (MS) patients in clinical trials (n = 755, 3269 scans) and from subjects participating in the Alzheimer's Disease Neuroimaging Initiative (ADNI, n = 834, 6114 scans). The percent change in BPF was modeled with a linear mixed effect (LME) model, and the model was applied separately to the MS and ADNI datasets. The LME model for the MS datasets included random subject effects (intercept and slope over time) and fixed effects for the time-of-day, time from the baseline scan, and trial, which accounted for trial-related effects (for example, different inclusion criteria and imaging protocol). The model for ADNI additionally included the demographics (baseline age, sex, subject type [normal, mild cognitive impairment, or Alzheimer's disease], and interaction between subject type and time from baseline).

There was a statistically significant effect of time-of-day on the BPF change in MS clinical trial datasets (−0.180 per day, that is, 0.180% of intracranial volume, $p = 0.019$) as well as the ADNI dataset (−0.438 per day, that is, 0.438% of intracranial volume, $p < 0.0001$), showing that the brain volume is greater in the morning. Linearly correcting the BPF values with the time-of-day reduced the required sample size to detect a 25% treatment effect (80% power and 0.05 significance level) on change in brain volume from 2 time-points over a period of 1 year by 2.6%.

Our results have significant implications for future brain volumetric studies, suggesting that there is a potential acquisition time bias that should be randomized or statistically controlled to account for the day-to-day brain volume fluctuations.

Introduction

Chronic brain atrophy has been well characterized pathologically in neurological diseases such as Alzheimer's disease (AD) (Alzheimer, 1907; Stelzmann et al., 1995) and multiple sclerosis (MS) (Dawson, 1916), and is traditionally considered a marker of neurodegeneration and chronic tissue loss. However, contemporary brain imaging and analysis techniques (Ashburner and Ridgway, 2012; Fox and Freeborough, 1997; Leung et al., 2012; Rudick et al., 1999; Smith et al., 2002) are so sensitive and precise that they are able to measure changes in brain volume on the order of 0.2–0.5% (Caramanos et al., 2010; Chard et al., 2002; Rudick et al., 1999; Smith et al., 2001). Such small changes, which are on the order of those expected from normal aging over 1 year (0.1–0.3% (Fisher et al., 2008; Fotenos et al., 2005; Scahill et al., 2003)), may result from physiological fluctuations, such as changes in brain hydration, for example. Studies that experimentally manipulated hydration status via overnight thirsting and subsequent drinking of water showed hydration-related changes in brain volume as large as 0.7% (Duning et al., 2005; Nakamura et al., 2014a). As such, the brain volume change may include significant physiological effects in addition to the chronic irreversible changes associated with neurodegeneration, and this may confound the measurement and interpretation of longitudinal brain volume changes. This has implications for the use of brain volume measurements to assess the response to disease modifying therapies.

In the current study, we were interested in potential diurnal fluctuations in brain volume in various populations. Rather than acquiring

☆ A portion of data used in the preparation of this article was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.usc.edu/ADNI). The ADNI investigators contributed to the design and implementation of ADNI and provided data but did not participate in analysis or writing of this report.

* Corresponding author at: 3801 University Street, WB319, Montreal Neurological Institute, Montreal, Quebec H3A 2B4, Canada. Fax: 1+514 398 2975.
E-mail address: knakamura@mrs.mni.mcgill.ca (K. Nakamura).

http://dx.doi.org/10.1016/j.neuroimage.2015.05.077
1053-8119/

Table 1
Description of datasets.

Dataset	Subject Group	# sites	# subjects	Total # MRIs	Max # MRI per subject	Age of inclusion criteria	Demographic or clinical covariates	Study duration
RESTORE	MS	31	171	1475	10	18–60	None	52 weeks
DEFINE	MS	84	584	1794	4	18–55	None	2 years
ADNI	AD	52	195	1074	10	55–90	Baseline age, sex, subject group [AD, MCI, or Normal]	4 years
	MCI	57	408	3247	14			
	Normal	55	231	1793	12			
	ADNI Total	58	834	6114	14			

The table shows the analyzed population after removing data acquired before 7 a.m. and after 7 p.m.
Abbreviation: AD = Alzheimer's disease, MCI = mild cognitive impairment, MS = multiple sclerosis.

MRIs in a controlled manner from selected subjects within a day, we chose to retrospectively analyze large MRI datasets. These datasets were longitudinal studies with repeated MRIs and uncontrolled time of acquisition. Thus, we (1) measured the BPF from large longitudinal MRI studies, (2) estimated the effect of time-of-day on brain volume using a linear mixed effect model in terms of percent change in BPF, and (3) measured the effect of correcting the time-of-day on the statistical power in terms of sample size.

Methods

Datasets

We used three datasets to test our hypothesis that there are diurnal changes in brain volume. Two datasets came from MS clinical trials and a third from the Alzheimer's Disease Neuroimaging Initiative (ADNI), enabling independent tests. The use of longitudinal data allowed measurements of percent change in BPF to reduce inter-subject variances. Each study is briefly described below while key information that is applicable in this study is summarized in Table 1.

The first MS dataset comes from the RESTORE[1] (NCT01071083) trial. The patients (n = 175) with relapsing–remitting MS (RRMS) who had been on natalizumab for at least one year were randomized to placebo, methylprednisolone, intramuscular interferon beta-1a, glatiramer acetate, or natalizumab for 6 months and re-started natalizumab after 6 months (Fox et al., 2014).

The second MS dataset was the DEFINE[2] (NCT00420212) trial where RRMS patients (n = 1237) were randomized to placebo, 240 mg oral dimethyl fumarate twice a day, or thrice a day (Gold et al., 2012).

In the RESTORE and DEFINE trials, we did not have access to the demographic and clinical data and could not account for age, sex, disease duration, therapy, lesion load, or measures of disability, all of which are known to have associations with brain volume (Fisher, 2011; Rudick et al., 2000; Sormani et al., 2014). Including these covariates would likely increase the statistical power, but we compensated with the large number of MRIs (over 3000 scans from more than 750 subjects).

We also used a subset of the ADNI dataset (NCT00106899) [Supplementary Material 1], which included elderly subjects who were cognitively normal as well as those with mild cognitive impairment (MCI) and those with AD. The ADNI subset studied consisted of data up to 4 years of follow-up from 1.5 T MRI scanners as of December 18, 2013. We excluded 23 images due to poor image quality such as incomplete brain coverage, subject motion, and aliasing artifacts.

We removed MRIs acquired before 7 a.m. and after 7 p.m. as they may represent atypical MRI sessions, and the number of samples outside this period was very small (Fig. 1). The number of individual MRI scans removed was 71 for RESTORE, and 132 for DEFINE, and 58 for ADNI leaving 1475, 1794 and 6114 MRI scans for statistical analysis, respectively. The numbers of sites, subjects, and MRIs used are summarized in Table 1.

Image analysis

T1-weighted images (acquisition parameters described in Table 2) were pre-processed. Our pre-processing for the MS data consisted of the following steps: (1) initial N3 correction with the distance option of 150 mm (Sled et al., 1998), (2) intensity de-noising using non-local means (Coupe et al., 2008), (3) 9 parameter (3 translations, 3 rotations, and 3 scalings) standard-space registration (Collins et al., 1994) using the ICBM 2009c nonlinear symmetric template[3] (Fonov et al., 2009), (4) intensity normalization with a single scaling factor, (5) brain extraction using BEaST (version 1.15)[4] (Eskildsen et al., 2012), (6) another application of N3 using the BEaST-derived brain extracted mask, (7) improved 12-parameter standard-space registration with the BEaST mask using the previous 9-parameter transformation as the initial estimate, and finally (8) MS white matter lesions were filled by the mean and standard deviation of local (4 mm radius) normal-appearing white matter tissues where normal-appearing white matter was obtained by applying FAST (FSL[5]) (Zhang et al., 2001) while excluding lesions. The spatial scaling in linear registration is used only for BEaST brain extraction. An example of original, pre-processed, and final images is shown in Fig. 2 for an MS case. The ADNI dataset was already pre-processed for B1 correction, N3 intensity nonuniformity correction, and geometric distortion. We did not perform any additional N3 corrections nor any lesion-inpainting for the ADNI dataset.

We developed our own implementation of BPF and did not use the original BPF by Fisher et al. (1997). BPF was more reproducible than the absolute brain volume (Fisher et al., 1997). Our BPF was calculated by [brain volume] / [intracranial volume]. The intracranial volume was estimated using BEaST (Eskildsen et al., 2012) as the number of voxels within the brain-extracted binary mask times the voxel size. The brain volume was calculated within the BEaST mask by automatic thresholding (Ridler and Calvard, 1978) after partial volume correction (Santago and Gage, 1995). The BEaST volume changes very slightly over time within individuals, but this will not likely bias the effect of time of day. The BPF values are not very sensitive to geometric distortion as the smoothly-varying distortion field affects both brain and intracranial volumes similarly. As in our previous studies (Nakamura et al., 2014b), MS white matter lesions were filled and did not influence BPF measurements. Other non-MS T1-hypointense white matter voxels were mostly retained within the brain volume after the thresholding procedure. Extremely hypointense white matter voxels were rare and not corrected. Examples of processed images from ADNI are shown in Fig. 3.

[1] Randomized treatment interruption of natalizumab.
[2] Randomized, multicenter, double-blind, placebo-controlled, dose-comparison study to determine the efficacy and safety of BG00012 in subjects with relapsing–remitting multiple sclerosis.

[3] http://www.bic.mni.mcgill.ca/ServicesAtlases/.
[4] http://www.bic.mni.mcgill.ca/ServicesSoftwareAdvancedImageProcessingTools/BEaST.
[5] http://fsl.fmrib.ox.ac.uk/.

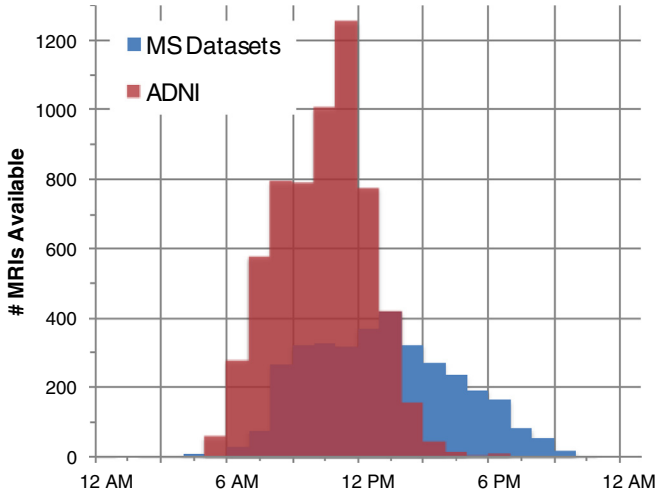

Fig. 1. Distribution of MRI acquisition times for the MS dataset (blue) and ADNI dataset (red) before removing scans acquired before 7 a.m. and after 7 p.m.

Statistical analysis

The time-of-day was determined from the acquisition time in the header of the Digital Imaging and Communications in Medicine (DICOM) file. As we are using multiple time-points per subject to model the effect of longitudinal volume change, the time from baseline in years (interval) was also calculated from the acquisition date in the DICOM header. To measure the extent of association between BPF and time-of-day, we applied the linear mixed effect (LME) model (Bates and Maechler, 2009) in the R statistical package (R-Team, 2012) for the MS datasets and separately for the ADNI dataset. The LME model is a powerful repeated-measures statistical model that contains both fixed and random effects.

In the first analysis, the change in BPF was modeled from MS patients in RESTORE and DEFINE using a single LME model. Random intercepts and slopes over the interval (time from baseline in years) were included for individual subjects. The fixed effects were the study ID (RESTORE or DEFINE), interval, and time-of-day. The study term accounted for the study-related effects such as different inclusion criteria and imaging protocols. The time-of-day variable was a continuous variable, and one unit of time-of-day was 24 hours. The model equation is shown in Eq. (1).

$$\Delta BPF_{i,j} = \beta_0 + \gamma_{0,i} + \left(\beta_1 + \gamma_{1,i}\right)[Interval] + \beta_2[TimeOfDay] + \beta_3[StudyID] + \varepsilon_{i,j} \quad (1)$$

where

- $\Delta BPF_{i,j}$ is the percent change in BPF from the first scan for the time-point j for subject i;
- [Interval] is the time between the baseline and follow-up scans in years;
- [TimeOfDay] is the time of day scaled from 0 to 1 for 24 hours;
- [StudyID] is RESTORE or DEFINE;
- $\beta_{0,1,2,3}$ are the fixed effect coefficients; β_0 is the intercept, β_1 is the linear slope for [Interval], β_2 is the linear slope for [TimeOfDay], β_3 is the coefficients for the binary variable [StudyID], that multiplies to 1 or 0 for RESTORE or DEFINE, respectively;
- $\gamma_{0,i}$ and $\gamma_{1,i}$ are the random effect coefficient for subject i and assumed to have zero mean and a constant variance;
- $\varepsilon_{i,j}$ is the error for time-point j for subject i.

The subject-specific effects in the mixed effect model imply a correlation between measurements from the same subject. The model estimates this correlation and accounts for multiple measurements taken from the same subject. The appropriate fixed effects are then added to the random effects to get a prediction of each BPF value. The residual error between the prediction and measurement is measured and used for significance calculations.

In the second analysis, the change in BPF was modeled from the ADNI dataset by the LME model. The model included random slopes and intercepts, and fixed effects for the interval, time-of-day, sex (male or female), interaction between the baseline age and interval, interaction between the subject type and interval, and interaction between the subject type and time-of-day and is shown in Eq. (2).

$$\Delta BPF = \beta_0 + \gamma_{0,i} + \left(\beta_1 + \gamma_{1,i}\right)[\text{Interval}] + \beta_2[\text{TimeOfDay}] + \beta_3[\text{BaselineAge} * \text{Interval}] + \beta_4[\text{SubjectType} * \text{Interval}] + \beta_5[\text{Sex} * \text{Interval}] + \beta_6[\text{SubjectType} * \text{TimeOfDay}] + \varepsilon_{i,j} \quad (2)$$

where the additional variables are explained below:

- [BaselineAge*Interval] is the interaction between the subject's age at the baseline scan and the interval;
- [SubjectType*Interval] is the interaction between the subject type (that is, Normal, MCI, or AD) and the interval; the subject type is a categorical variable, which consists of 3 hidden binary variables, and each binary variable multiplies 1 or 0 for Normal or non-Normal, 1 or 0 for MCI or non-MCI, and 1 or 0 for AD or non-AD;
- [Sex*Interval] is the interaction between the subject's sex and interval where sex is a binary variable that multiplies 1 or 0 for Male or Female;
- [SubjectType*TimeOfDay] is the interaction between the subject type and the time-of-day;
- $\beta_{0,\ldots,6}$ are the fixed effect coefficients as before; β_0 is the intercept, β_1 is the linear slope for [Interval], β_2 is the linear slope for [TimeOfDay], β_3 is the coefficient for the interaction term [BaselineAge*Interval], β_4 is the coefficient for the interaction term [SubjectType*Interval]; β_5 is the coefficient for the interaction term [Sex*Interval]; β_6 is the coefficient for the interaction term [SubjectType*TimeOfDay].

Table 2
MRI parameters for each study.

Study	Modality	Echo time (ms)	Repetition time (ms)	Flip angle (degree)	Resolution (mm^3)	Field strength (# sites)	Manufacturer (# sites)
RESTORE	Axial 3D T1-weighted gradient echo (FLASH)	8–12 (1.5 T) 4–6 (3 T)	30 (1.5 T) 25 (3 T)	30	1.0 × 1.0 × 3.0	1.5 T (24) 3.0 T (7)	GE (17) Hitachi (1) Philips (4) Siemens (9)
DEFINE	Axial 3D T1-weighted gradient echo (FLASH)	8–11 (1.5 T) 5 (3 T)	30–35 (1.5 T) 26 (3 T)	30	1.0 × 1.0 × 3.0	1.0 T (6) 1.5 T (76) 3.0 T (1) Unknown (1)	GE (28) Marconi (2) Philips (15) Picker (2)Siemens (36) Toshiba (1)
ADNI	Sagittal 3D T1-weighted MPRAGE	3.5–4.1	8.55–10.40	8	(0.9375–1.25) × (0.9375–1.25) × 1.2	1.5 T (58)	GE (27) Philips (9) Siemens (22)

Abbreviations: FLASH = Fast Low Angle SHot, MPRAGE = magnetization prepared rapid acquisition with gradient echo, GE = General Electric.

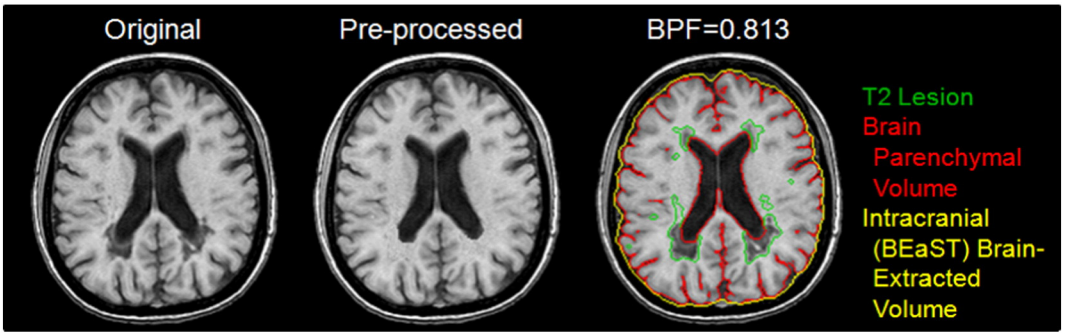

Fig. 2. Example of image processing from a multiple sclerosis (MS) patient. The original T1-weighted image (left) is pre-processed including T2 lesion inpainting (middle). The color image (right) represents semi-automatically segmented MS T2-lesion (green contour), the BEaST brain extracted mask (yellow), and the brain parenchymal volume (red). The brain parenchymal fraction (BPF) for this case was 0.813.

To estimate the effect of correcting the time-of-day on the statistical power, we calculated the required sample sizes to detect significant differences in brain atrophy rates in a hypothetical 1-year longitudinal AD study with 2 time-points. We used a subset of the ADNI data consisting of pairs of MRIs acquired at baseline and 1-year from 190 normal controls and 130 patients with AD. The sample size was calculated by a standard equation in the literature (Fitzmaurice et al., 2004; Fox et al., 2000):

$$N = Ceil\left(\frac{2(SD_p)^2(u+v)^2}{([RxE](\mu_{AD}-\mu_{Normal}))^2}\right) \quad (3)$$

where N is the required sample size per arm, *Ceil* is a round-up operator, $u = 0.842$ for 80% power, and $v = 1.96$ for the 5% significance level. μ_{AD} and μ_{Normal} are the mean rates of brain volume change for AD and normal groups, respectively, thus accounting for normal aging. SD_p is a pooled standard deviation where $SD_p = \sqrt{\frac{(n_{AD}-1)SD_{AD}^2+(n_{Normal}-1)SD_{Normal}^2}{n_{AD}+n_{Normal}-2}}$, where n is the number of subjects and SD is the standard deviation for each group. RxE is the predefined treatment effect (25%), which is assumed to take immediate and constant effect. The 25% treatment effect was chosen from the commonly used values in the literature (Holland et al., 2009; McEvoy et al., 2010). We calculated the sample size per arm from the unadjusted BPF and then from the unadjusted BPF values corrected for the time-of-day according to the fixed effect term obtained from the LME model in the entire ADNI dataset.

The MS datasets included treated and non-treated patients, and we remain blinded to treatment status. Therefore, we could not perform proper sample size estimation for the MS datasets since the variance estimates for atrophy would be overestimated due to treatment heterogeneity. Instead, we report the percent difference in the sample size estimation before and after the time-of-day correction without accounting for normal aging. We used the 1-year data from the DEFINE dataset. The equation was slightly modified as below:

$$N = Ceil\left(\frac{2(SD)^2(u+v)^2}{([RxE]\mu)^2}\right). \quad (4)$$

We obtained the required sample size per arm with or without time-of-day correction and calculated the percent change in the sample sizes.

Results

The histograms of measured BPF values are plotted for the MS and ADNI group in Fig. 4. The unadjusted mean (SD) BPF values were 0.876 (0.030) for MS and 0.814 (0.027) for ADNI. The correlation between the time-of-day and time from baseline in years (interval) was low for both datasets (0.060 for MS and 0.003 for ADNI).

From the MS clinical trials, the fixed effect of time-of-day on BPF was significant (-0.180% per day, standard error (SE) $= 0.076$, $p = 0.0188$); also significant were the time from baseline (-0.253% per year, $p < 0.0001$) and study (compared to DEFINE, the effect for RESTORE was $+0.072\%$, $p = 0.0090$). Fig. 5 shows the plot of model-corrected BPF and time-of-day.

From ADNI (Fig. 6), the interaction between the subject type and time-of-day was not significant ($p = 0.2438$). The interaction between the interval and sex was also not significant ($p = 0.2955$), and these terms were removed from subsequent analysis. Significant effects

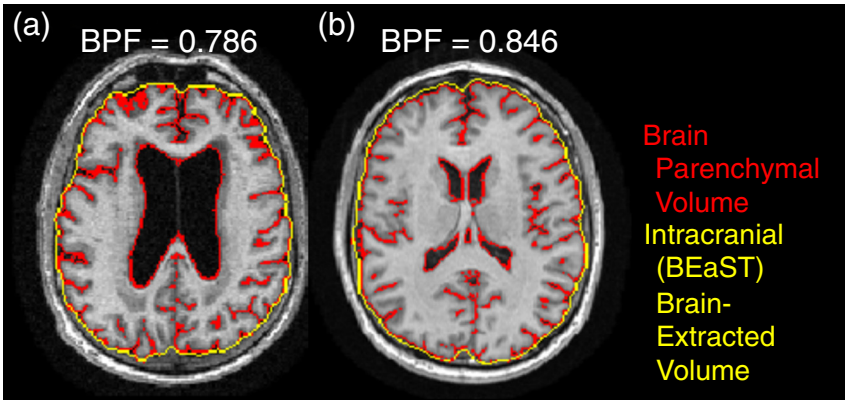

Fig. 3. Example segmentations from ADNI data. The BEaST brain extracted mask is outlined in yellow, and the brain parenchymal volume is outlined in red. (a) is from a subject with Alzheimer's disease. Hypointense areas in posterior and anterior periventricular white matter are classified as brain parenchyma. (b) is from a normal subject. The brain parenchymal fraction (BPF) was 0.786 for (a) and 0.846 for (b).

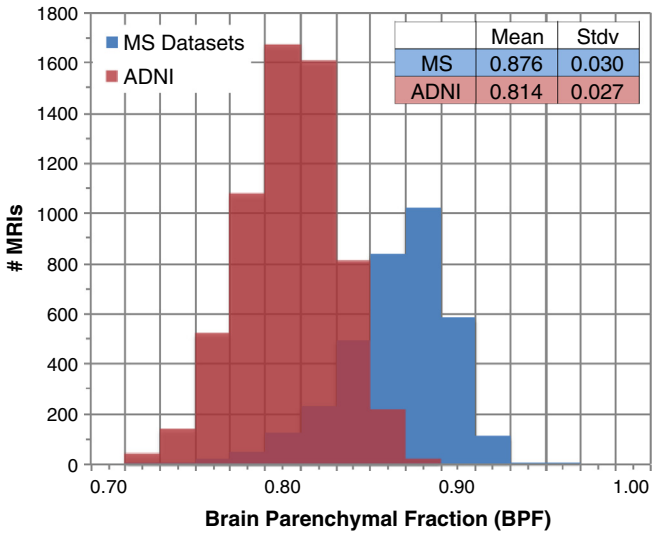

Fig. 4. Distribution of unadjusted brain parenchymal fraction (BPF) for the MS datasets (blue) and ADNI dataset (red). The overall mean ± standard deviation BPF values were 0.876 ± 0.030 for MS datasets and 0.814 ± 0.027 for ADNI.

were the following: time-of-day (−0.442% per day, SE = 0.071, $p < 0.0001$), interval (time from baseline, −0.913% per year, $p < 0.0001$), baseline age (0.006% per year, $p < 0.0107$), and subject type (compared to normal, AD = −0.673% per year, MCI = −0.336% per year, $p < 0.0001$).

Finally, the sample size per arm to detect a 25% treatment effect on brain atrophy in AD trial using BPF was 272 for 80% power and 0.05-significance level in a hypothetical 1-year study with 2 time-points after accounting for normal aging. When corrected for the effect of time-of-day (−0.221% per 12 hours), the sample size decreased by 2.6% to 265. For the MS population, the percent reduction in the required sample size after correcting for the time-of-day was 2.8%.

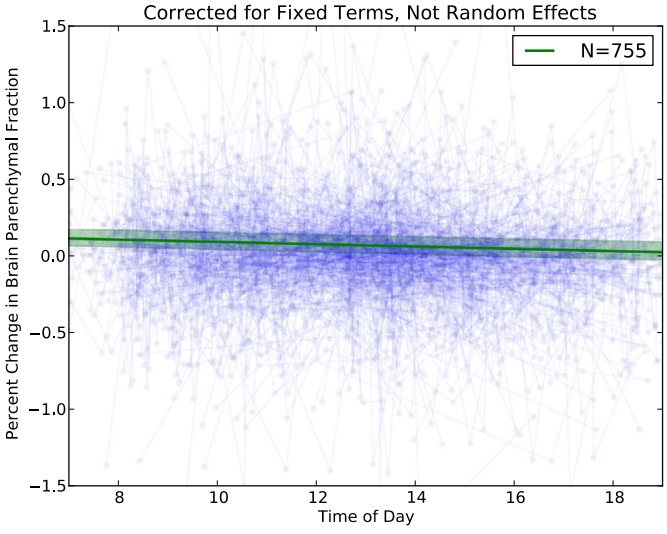

Fig. 5. Model-corrected change in brain parenchymal fraction (BPF) plotted against the time-of-day using 3269 MRIs of multiple sclerosis patients (n = 755). The BPF was corrected for all fixed effects except the time-of-day; that is, the time from baseline in years (interval) and study (DEFINE or RESTORE). There was a statistically significant effect of time-of-day on change in BPF (−0.090% per 12 hours, $p = 0.0188$). The green line indicates the model estimate of the fixed effect of time of day, the light green region surrounding the green line shows the 95% confidence interval on the fitted model. The blue circles indicate individual measurements, and connected lines indicate subjects.

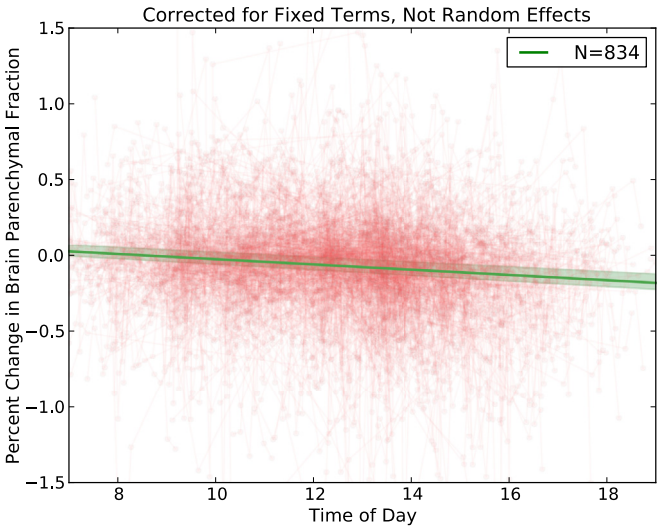

Fig. 6. Model-corrected change in brain parenchymal fraction (BPF) plotted against the time-of-day from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (6114 scans from 834 subjects). The change in BPF was adjusted for all fixed effects except for the time-of-day; that is, the time from baseline in years (interval), interaction between the baseline age and interval, interaction between the subject type (AD, MCI, or Normal) and interval. There was a statistically significant effect of time-of-day on change in BPF (−0.221% per 12 hours, $p < 0.0001$). The green line indicates the model estimate of the fixed effect of time of day and the light green region surrounding the green line shows the 95% confidence interval on the fitted model. The red circles indicate individual measurements, and connected lines indicate subjects.

Discussion

We found a statistically significant effect of time-of-day on the change in BPF in two different populations—MS subjects and elderly subjects with or without dementia. The mean effects were different (−0.090% and −0.221% per 12 hours) in these groups, but they both showed a significant effect of time-of-day on the change in BPF. Furthermore, the group characteristics also were different: age range (maximum possible age of 60 in the MS studies and mean baseline age of 75.3 in ADNI), imaging protocols (standard-resolution FLASH in MS and high-resolution MPRAGE in ADNI), availability of demographic information, the number of repeated scans (on average, 4.3 scans per subject for MS and 7.3 scans per subject for ADNI), and the image processing pipelines not identical. Despite these differences, significant decreases in brain volume were found in both datasets.

The results show that subjects' brains are larger in the morning. A possible mechanism may be that lying down during the night is associated with a redistribution of body fluids that had accumulated in the lower extremities during the day. The effect of time-of-day (−0.090 and −0.221% per 12 hours) is similar to the yearly rate of brain atrophy in MS (−0.253% per year) and elderly normal subjects (−0.913% per year). It is inconceivable that the brain volume persistently decreases by such amount every day. Therefore, this volume decrease must be accompanied by a similar volume increase overnight.

It is also possible that the effect of time-of-day is associated with hydration status. The imaging protocols did not include T2 relaxation time, which would have allowed measurement of brain water content (Whittall et al., 1997). Examples of mechanisms that act through hydration status include electrolytes and osmolytes (Gullans and Verbalis, 1993) as well as cortisol, which is diuretic and elevated in the morning. We also cannot exclude the possibility that the daily medication schedule, some of which are diuretic, may affect the brain volume. Although our study design is retrospective and such data are not available, the similar pattern of brain volume change throughout the day in various populations (MS, elderly normal, MCI and AD) suggests that our finding is genuine and general across populations.

It is also possible that a systematic imaging bias may play a role; for example, the gradient coil may not have warmed up in the first morning scan. But, this is unlikely, as other MRI contrasts (such as PD-weighted or T2-weighted images) were acquired before the T1-weighted image, at least in the MS datasets. Another example is the scan order, which is typically consistent within session and has not been fully addressed in the current study.

As a post-hoc analysis, we investigated a potential correlation between time-of-day and covariates by testing for the associations between acquisition time and demographic variables in the ADNI datasets using LME models. There were no significant differences in acquisition time between the groups ($p > 0.36$), with sex ($p = 0.09$), with subject age ($p = 0.14$), or with interval ($p = 0.53$), supporting the validity of our study.

We used a variation of BPF, and our BPF values are slightly higher than those obtained using the original BPF method (Fisher et al., 1997; Nakamura et al., 2010; Rudick et al., 1999), but the brain atrophy rates are in line with previous studies of MS using the original BPF (Fisher et al., 2008; Miller et al., 2007). The BEaST brain extraction mask provided an approximation of cranial volume, as it excluded the intracranial veins. The BEaST volume changes slightly over time within individuals (data not shown), but the changes are an order of magnitude smaller than the brain volume change (data not shown). Thus, the use of BEaST mask may slightly under-estimate the longitudinal brain atrophy, but that is unlikely to bias the effect of time of day, as the scan times were random.

The effect on sample size of correcting for the time-of-day was small, with only 2.6% reduction despite the large and significant effect. Varying the treatment effect between 5 and 95% did not significantly change this relative reduction (0–4.8%). The large variance observed on Figs. 5 and 6 in the time-of-day effect explains this small improvement; the time-of-day does not account for all physiologic factors. The low correlation between corrected BPF and time-of-day (Figs. 5 and 6) corroborates this explanation.

Despite the small contribution to the sample size estimated from the *large* ADNI and MS datasets, our results have great implications on future brain atrophy studies. Our study suggests that a bias related to the acquisition time exists, and this may be especially apparent in small studies where the time of image acquisition may not be fully random. For example, a tendency to acquire MRIs from healthy subjects in the morning and the diseased group in the afternoon would bias the brain volumes towards a greater group difference in cross-sectional studies. Similarly, acquiring baseline images in the morning and follow-up images in the afternoon would cause artificially accelerated brain atrophy in longitudinal studies.

Based on our results, we recommend to control for the time-of-day in small studies where the scan times may not be fully random, by chance or by operational constraints (for example, hours of operation in clinical/research scanners and bias in visit schedule of different subject groups), to reduce this potential effect. If the effect of time-of-day on the measurement can be well established, statistically accounting for the effect of time-of-day may also suffice.

In conclusion, we have identified a major effect of normal reversible brain volume fluctuation, which could have significant impact on volumetric MRI studies of brain.

Acknowledgments

The authors thank Biogen Idec for the use of MRI data from Randomized Treatment Interruption of Natalizumab (RESTORE, NCT01071083) and data from Randomized, Multicenter, Double-Blind, Placebo-Controlled, Dose-Comparison Study to Determine the Efficacy and Safety of BG00012 in Subjects With Relapsing–Remitting Multiple Sclerosis (DEFINE, NCT00420212) trial; and NeuroRx Research for providing these data.

Portions of data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Rev December 5, 2013 Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

K. Nakamura is supported by the post-graduate award from Mitacs Elevate Postdoctoral Fellowship.

References

Alzheimer, A., 1907. Uber eine eigenartige Erkrankung der Hirnrinde. Allg. Z. Psychiatr. Phys. Gerichtl. Med. 64, 146–148.
Ashburner, J., Ridgway, G.R., 2012. Symmetric diffeomorphic modeling of longitudinal structural MRI. Front. Neurosci. 6, 197.
Bates, D., Maechler, M., 2009. Ime4: Linear Mixed-Effects Models Using S4 Classes R Package Version 0.999375-32.
Caramanos, Z., Fonov, V.S., Francis, S.J., Narayanan, S., Pike, G.B., Collins, D.L., Arnold, D.L., 2010. Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. NeuroImage 49, 1601–1611.
Chard, D.T., Parker, G.J., Griffin, C., Thompson, A.J., Miller, D.H., 2002. The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology. J. Magn. Reson. Imaging 15, 259–267.
Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.
Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C., 2008. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans. Med. Imaging 27, 425–441.
Dawson, J., 1916. The histology of multiple sclerosis. Trans. R. Soc. Edinb. 50, 517–740.
Duning, T., Kloska, S., Steinstrater, O., Kugel, H., Heindel, W., Knecht, S., 2005. Dehydration confounds the assessment of brain atrophy. Neurology 64, 548–550.
Eskildsen, S.F., Coupe, P., Fonov, V., Manjon, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Ostergaard, L.R., Collins, D.L., 2012. BEaST: brain extraction based on nonlocal segmentation technique. NeuroImage 59, 2362–2373.
Fisher, E., 2011. Measurement of CNS atrophy. In: Cohen, J.A., Rudick, R.A. (Eds.), Multiple Sclerosis Therapeutics. Cambridge University Press, pp. 128–149.
Fisher, E., Cothren, J.R.M., Tkach, J.A., Masaryk, T.J., Cornhill, J.F., 1997. Knowledge-based 3D segmentation of the brain in MR images for quantitative multiple sclerosis lesion tracking. Proc. SPIE 3034, 19–25 (Medical Imaging).
Fisher, E., Lee, J.C., Nakamura, K., Rudick, R.A., 2008. Gray matter atrophy in multiple sclerosis: a longitudinal study. Ann. Neurol. 64, 255–265.
Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2004. Applied Longitudinal Analysis. Wiley-Interscience, Hoboken, N.J.
Fonov, V., Evans, A., McKinstry, R., Almli, C., Collins, D., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage 47, S102.
Fotenos, A.F., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L., 2005. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology 64, 1032–1039.
Fox, N.C., Freeborough, P.A., 1997. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. J. Magn. Reson. Imaging 7, 1069–1075.
Fox, N.C., Cousens, S., Scahill, R., Harvey, R.J., Rossor, M.N., 2000. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. Arch. Neurol. 57, 339–344.
Fox, R.J., Campbell Cree, B.A., De Seze, J., Gold, R., Hartung, H.P., Jeffery, D., Kappos, L., Kaufman, M., Montalban, X., Weinstock-Guttman, B., Anderson, B., Natarajan, A.,

Ticho, B., Duda, P., 2014. MS disease activity in RESTORE: a randomized 24-week natalizumab treatment interruption study. Neurology 82, 1491–1498.
Gold, R., Kappos, L., Arnold, D.L., Bar-Or, A., Giovannoni, G., Selmaj, K., Tornatore, C., Sweetser, M.T., Yang, M., Sheikh, S.I., Dawson, K.T., Investigators, D.S., 2012. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. N. Engl. J. Med. 367, 1098–1107.
Gullans, S.R., Verbalis, J.G., 1993. Control of brain volume during hyperosmolar and hypoosmolar conditions. Annu. Rev. Med. 44, 289–301.
Holland, D., Brewer, J.B., Hagler, D.J., Fennema-Notestine, C., Dale, A.M., Weiner, M., Thal, L., Petersen, R., Jack, C.R., Jagust, W., 2009. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proc. Natl. Acad. Sci. 106, 20954–20959.
Leung, K.K., Ridgway, G.R., Ourselin, S., Fox, N.C., 2012. Consistent multi-time-point brain atrophy estimation from the boundary shift integral. NeuroImage 59, 3995–4005.
McEvoy, L.K., Edland, S.D., Holland, D., Hagler Jr., D.J., Roddey, J.C., Fennema-Notestine, C., Salmon, D.P., Koyama, A.K., Aisen, P.S., Brewer, J.B., 2010. Neuroimaging enrichment strategy for secondary prevention trials in Alzheimer's disease. Alzheimer Dis. Assoc. Disord. 24, 269.
Miller, D.H., Soon, D., Fernando, K.T., MacManus, D.G., Barker, G.J., Yousry, T.A., Fisher, E., O'Connor, P.W., Phillips, J.T., Polman, C.H., Kappos, L., Hutchinson, M., Havrdova, E., Lublin, F.D., Giovannoni, G., Wajgt, A., Rudick, R., Lynn, F., Panzara, M.A., Sandrock, A.W., Investigators, Affirm, 2007. MRI outcomes in a placebo-controlled trial of natalizumab in relapsing MS. Neurology 68, 1390–1401.
Nakamura, K., Rudick, R.A., Lee, J.C., Foulds, P., Fisher, E., 2010. Effect of Intramuscular Interferon beta-1a on Gray Matter Atrophy in Relapsing–Remitting Multiple Sclerosis. American Academy of Neurology (AAN), Toronto, Ontario, Canada.
Nakamura, K., Brown, R.A., Araujo, D., Narayanan, S., Arnold, D.L., 2014a. Correlation between brain volume change and T2 relaxation time induced by dehydration and rehydration: implications for monitoring atrophy in clinical studies. NeuroImage Clin. 6, 166–170.
Nakamura, K., Guizard, N., Fonov, V.S., Narayanan, S., Collins, D.L., Arnold, D.L., 2014b. Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis. NeuroImage Clin. 4, 10–17.
Ridler, T., Calvard, S., 1978. Picture thresholding using an iterative selection method. IEEE Trans. Syst. Man Cybern. 8, 630–632.
R-Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 3-900051-07-0 (2007).
Rudick, R., Fisher, E., Lee, J.-C., Simon, J., Jacobs, L., 1999. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing–remitting MS. Neurology 53, 1698.
Rudick, R.A., Fisher, E., Lee, J.C., Duda, J.T., Simon, J., 2000. Brain atrophy in relapsing multiple sclerosis: relationship to relapses, EDSS, and treatment with interferon beta-1a. Mult. Scler. 6, 365–372.
Santago, P., Gage, H.D., 1995. Statistical-models of partial volume effect. IEEE Trans. Image Process. 4, 1531–1540.
Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., Fox, N.C., 2003. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. Arch. Neurol. 60, 989–994.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. Med. Imaging IEEE Trans. 17, 87–97.
Smith, S.M., De Stefano, N., Jenkinson, M., Matthews, P.M., 2001. Normalized accurate measurement of longitudinal brain change. J. Comput. Assist. Tomogr. 25, 466–475.
Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17, 479–489.
Sormani, M.P., Arnold, D.L., De Stefano, N., 2014. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. Ann. Neurol. 75, 43–49.
Stelzmann, R.A., Norman Schnitzlein, H., Reed Murtagh, F., 1995. An English translation of Alzheimer's 1907 paper, "Über eine eigenartige Erkankung der Hirnrinde". Clin. Anat. 8, 429–431.
Whittall, K.P., MacKay, A.L., Graeb, D.A., Nugent, R.A., Li, D.K., Paty, D.W., 1997. In vivo measurement of T2 distributions and water contents in normal human brain. Magn. Reson. Med. 37, 34–43.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.