Research Article   **Open Access**

Differentiating between Normal Aging, Mild Cognitive Impairment, and Alzheimer's disease with FDG-PET: Effects of Normalization Region and Partial Volume Correction Method

Corinna M Bauer[1], Howard J Cabral[2], Douglas N Greve[3], Ronald J. Killiany[4,5,6]

[1]Boston University School of Medicine, Department of Anatomy and Neurobiology, Boston, MA, USA
[2]Boston University School of Public Health, Department of Biostatistics, Massachusetts Avenue, Boston, MA, USA
[3]Martinos Center for Biomedical Imaging, 143 13th Street, Charlestown, MA, USA
[4]Director, Center for Biomedical Imaging, Boston University School of Medicine, Boston, Massachusetts, USA
[5]Associate Professor, Department of Anatomy and Neurobiology, Boston University, School of Medicine, Boston, Massachusetts, USA
[6]Associate Professor, Department of Environmental Health, Boston University School of Public Health, Boston, Massachusetts, USA

Abstract

Objective: In Alzheimer's FDG PET research, the choice of reference region for normalization and use of partial volume correction are inconsistent and have not been studied in a large multi-center study. Herein, we identified which normalization region provided the highest degree of discrimination between subjects who were classified as normal aging, mild cognitive impairment, or Alzheimer's disease. The effects of partial volume correction using either a gray matter mask or cortical thickness and subcortical volume residuals were also examined.

Methods: Stepwise logistic regression models were used to identify the optimal normalization region and partial volume correction method to discriminate between disease stages in over 400 subjects from research sites across North America. Normalization region candidates were the brainstem, precentral gyrus, postcentral gyrus, cerebellum, and thalamus. Partial volume correction methods tested were anatomically or statistically based.

Results: Pre- and post- central gyri, and the thalamus showed AD-related changes in FDG PET and did not qualify for further testing. Normalizing to the cerebellum while using the gray matter mask for partial volume correction provided the highest indicator of discrimination.

Conclusions: Normalization region and partial volume correction are critical to FDG PET analysis and candidate normalization regions should be tested for disease effects in the study sample prior to use. Cerebellar normalization and gray matter mask partial volume correction are recommended for use with the ADNI dataset.

Keywords: FDG PET; Partial Volume Correction; Mild cognitive impairment; Alzheimer's disease; Normalization region

Introduction

Alzheimer's disease affects an estimated 36 million people worldwide and the incidence of this disease is projected to increase substantially as the global population continues to age. The prevalence of mild cognitive impairment (MCI), a precursor to AD, is also on the rise and treatment for this condition is hampered by our inability to accurately diagnose this condition. Thus, there is a pressing need to identify markers for those who are aging normally, have MCI, or have AD. One potential candidate is decreased glucose consumption in the brain which has been correlated with memory function in MCI and AD, as measured with ^{18}F-flurodeoxyglucose (FDG) PET [1]. However the accuracy of FDG PET metabolic values can be heavily influenced by the post-processing methods employed.

Normalization is a post-processing step that utilizes a reference region to standardize the data and allows for the direct comparison of non-quantitative PET data. A candidate reference region for normalization should not be affected by the disease or condition being studied, be easily and reliably identifiable, and ought to provide the most accurate differentiation of the groups [2,3]. The optimal choice of reference region for normalization has been debated in the field, with the pons [4], cerebellum [2,5-7], and primary somatosensory cortices [3] all being suggested to be the most appropriate region to use. However, each of these studies is limited in sample size and to one research center, thus the results may not be widely applicable.

Another critical step in the post-processing FDG PET stream

is partial volume correction. This is particularly important in MCI and AD research because disease-related tissue atrophy can result in artificially low uptake values thereby falsely enhancing any difference between the disease and control groups [8]. If partial volume correction is not used, then any observed hypometabolism is likely contaminated by tissue atrophy. To eliminate this problem, some FDG PET studies use partial volume correction (PVC) in a voxel-wise [9-14] or a region of interest (ROI)-based approach [15]; however, which PVC method enables the best ability to differentiate between normal aging, MCI, and AD subject groups has not been examined in a large sample.

The main aims of this study were to first examine which ROI provides the most sensitive normalization when differentiating between normal aging, MCI, and AD in a large sample collected at multiple research centers. The candidate regions, cerebellum, thalamus, brainstem, and pre- and post- central gyri, have previously shown relatively preserved glucose metabolism [16] and all, with the exception of the thalamus,

have previously held claim to the best region for normalization. The second aim of this study was to examine the effects of PVC on normal aging, MCI, and AD using two techniques: 1) applying a grey matter mask (GMM) [14,17] and 2) taking residuals of uptake after removing the effects of cortical thickness plus subcortical volume [15]. This is the first study to our knowledge to examine these post-processing factors in a large dataset encompassing multiple research centers.

Materials and Methods

Subjects

The data for use in this study were chosen from the larger pool of data that has been made publically available by the Alzheimer's Disease Neuroimaging Initiative. Data was screened to include all subjects who had both PET and MRI scans available for use on the ADNI/LONI website (www.loni.ucla.edu/ADNI) at the time this study began. From this screened dataset, PET data from 21 subjects was of poor contrast and quality and had to be omitted from the analyses for this study. Three subjects were omitted due to missing information. This left us with data from 403 subjects. Demographic information can be found in Table 1.

All subjects completed a battery of neuropsychological tests, including the mini-mental state examination (MMSE) [18], the CDR-Sum of Boxes [19], and the Global dementia scale (GD-scale) [20,21]. On the basis of their cognitive status the subjects were classified by the ADNI clinical core as: (a) normal controls with normal cognition and memory, CDR 0, and MMSE between 24-30; (b) amnestic MCI with memory complaint verified by a study partner, memory loss measured by education-adjusted performance on the Logical Memory II subscale of the Wechsler Memory Scale-Revised [22], preserved activities of daily living, CDR 0.5, MMSE between 24 and 30, and absence of dementia at time of baseline MRI scan; or (c) probable AD with memory complaint validated by an informant, abnormal memory function for age and education level, absence of depression, impaired activities of daily living, diminished cognition, CDR > 0.5, and MMSE between 20-26.

Alzheimer's Disease Neuroimaging Initiative

The ADNI was a 5-year non-randomized natural history non-treatment study utilizing data from multiple study centers across the United States and Canada. One of the main goals of the ADNI was to develop optimized methods and uniform standards for the acquisition of multicenter MRI and PET data on normal control subjects and patients with MCI and AD in drug/treatment trials. For more information about the ADNI please refer to http://www.adni-info.org.

MRI scans

For this study, we chose MRI scans from those acquired by the ADNI on 1.5T scanners from General Electric (GE), Philips Medical Systems (Philips), and Siemens Medical Solutions (Siemens). Specific pulse sequence guidelines can be found at http://www.loni.ucla.edu/ADNI/Research/Cores/index.shtml. In this study we used the two MPRAGE scan acquired at baseline for each participant. The data from the LONI website was downloaded for use in its original format since

Figure 1: Coregistration of FDGPET and T1-weighted MPRAGE MRI scans shown in coronal, sagittal, and axial sections. The green line indicates the gray/white matter boundary. The FDG PET scans are on the top and the MRI scans are on the bottom. The images are from a normal aging 83-year old male.
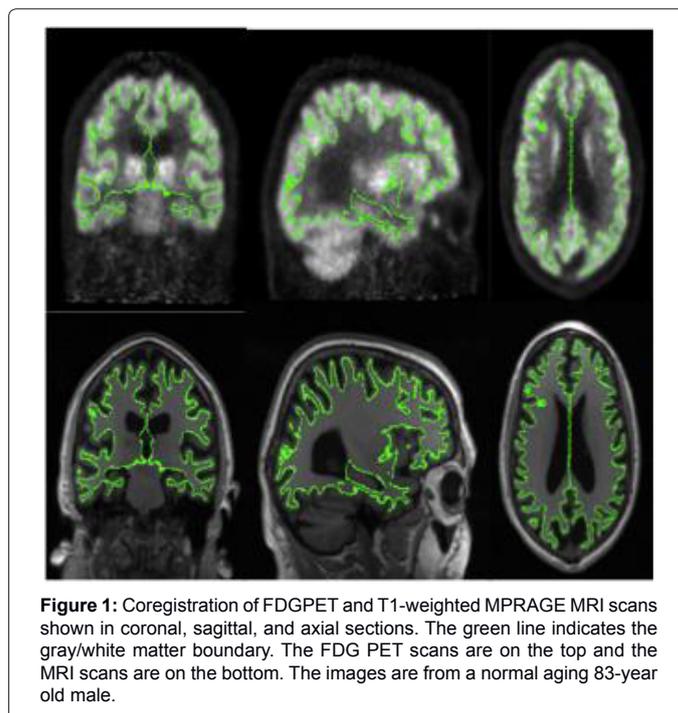
the Freesurfer processing pipeline has its own normalizing procedures.

FDG PET scans

For this study, we chose FDG PET scans from those acquired by the ADNI on GE, Philips, or Siemens scanners. Specific protocols for each scanner are available from the ADNI website (http://adni.loni.ucla.edu/research/protocols/pet-protocols/). These data were corrected for radiation attenuation and scatter using scanner-specific algorithms and each image was visually assessed for potential artifacts by the ADNI PET core at the University of Michigan. For this study we used the original PET data that was not pre-processed by the ADNI PET core so that we could have local control of all the processing steps as with the MRI scans.

Freesurfer Analysis

All MRI and PET scans were processed with the Freesurfer 4.4.1 image analysis suite, which is documented and freely available for download from http://surfer.nmr.mgh.harvard.edu/. The FDG PET scans were re-processed in version 5.1.0 in order to do GMM partial volume correction. For each subject, the 2 DICOM T1-weighted MRI datasets were motion corrected, averaged, segmented into gray matter, white matter, and cerebral spinal fluid (CSF), and intensity normalized. The cortex was parcellated into regions of interest based on gyral and sulcal structure. For each of the cortical regions volume, surface area, and cortical thickness were determined. Volume was calculated for each of the subcortical structures. Please refer to the FreeSurfer wiki page for more detailed information (http://surfer.nmr.mgh.harvard.edu/fswiki).

PET Processing

Once the T1-weighted MRI images were processed, the PET images were affine spatial transformed into "anatomical space" 1x1x1, 256x256x256, which was the same resolution as the transformed MRI images. The PET and MRI images were then co-registered using an automated Freesurfer boundary based application using 6 degrees of freedom, such that no skewing or twisting of the data occurred (Figure 1).

	Subjects (male/female)	Education Mean (std dev)	MMSE
Normal Aging	105 (64/41)	15.90 (3.12)	28.98 (1.12)
MCI	204 (137/67)	15.80 (2.88)	27.15 (1.71)[a]
AD	94 (56/38)	14.61 (3.21)[a,b]	23.48 (2.14)[a,b]

[a]significant difference from normal aging (p < 0.05), [b] significant difference from MCI (p< 0.05).\

Table 1: Demographic information.

The resulting coregistration was visually assessed for accuracy and adjusted if necessary (approximately 25% of the datasets). After the two datasets were co-registered, the PET data was applied as a mask to the MRI images and analyses performed. FDG uptake was measured in specific ROIs according to the cerebral cortex parcellations generated on the representative MRI images [23]. A total of 41 cortical and subcortical areas were examined for changes in MRI morphometry and FDG uptake related to MCI and AD relative to normal aging.

PET reference region for normalization

To control for individual global variations and to increase sensitivity of the method for differentiating between groups [3], a number of reference regions for normalization were tested, namely the brainstem, precentral gyrus, postcentral gyrus, cerebellum, and thalamus. The reference regions were first tested for disease effects before and after PVC and if there were no disease-related differences found using ANOVA, they were assessed as normalization regions. Normalization was done using residuals.

PET Gray matter mask partial volume correction

Each participant's MRI was segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid space (CSF). The MRI WM segment was made binary and, after registration, served as a WM mask for the PET images. The WM PET mask was then made binary and multiplied by the original PET image to obtain a WM-only PET image. The GM PET image with partial volume correction (PVC) was obtained by subtracting the WM PET image from the original PET image and then multiplying this image by a binary GM MRI mask. The Freesurfer parcellations were then applied to the GM PET.

PET residuals partial volume correction

The second PVC method was a statistical-based approach in which the effects of cortical thickness and subcortical volume were removed from the FDG PET metabolic values via residuals. This approach has been used in the past with FDG PET data processed in Free Surfer [15].

Statistical Analysis

In order to assess the equity of the male-female distribution in the three diagnostic groups, Chi square tests were performed. ANOVA was used in order to assess the age distribution in the three diagnostic groups. T-tests showed no significant left/right differences, thus the data from the two hemispheres was averaged.

Logistic regression Analyses

For each candidate normalization region we created a logistic stepwise regression model for diagnostic group, controlling for age, gender, and education by forcing them into the model. This was accomplished by running linear regressions for MCI vs. normal, AD vs. MCI, and all three groups together with each FDG PET regional uptake variable individually in order to identify their predictor values. If the point estimate was below 0.75 or above 1.25 the FDG PET regional uptake variable was determined to be an adequate predictor on its own and was entered as a variable in the overall model.

Separate models were created for differentiating between AD and MCI, MCI and normal aging, and all three groups together based upon data normalized to each of the candidate normalization regions. The c-statistic and AIC for each model were used to assess its utility. Hosmer-Lemeshow goodness of fit Chi square tests were used to assess the models' calibration.

Results

Chi-square tests revealed no significant differences for distribution of males and females between groups (df=2, p=0.3517). Age was not significantly different between control, MCI, and AD groups, as indicated with ANOVA (p=0.6684). Cluster analysis showed no significant grouping according to site, thus data from all research sites were grouped together.

Normalization region

We tested the pre- and post-central gyri, thalamus, brainstem, and cerebellum both before and after partial volume correction for significant differences between groups in order to assess if disease state has an effect on them. After these initial tests, only uptake in the brainstem and cerebellum was preserved as normalization candidates. ANOVA indicated that in the non-partial volume corrected PET data, the pre- and post-central gyri showed significantly decreased FDG uptake (p < 0.05) with disease progression eliminating them as potential normalization candidates. After partial volume correction, the thalamus showed significantly decreased uptake with disease (p < 0.05) and was eliminated as a potential candidate for normalization. Conversely, the brainstem and cerebellum did not show significant disease-related changes (p > 0.05) making them good candidates for normalizing the FDG PET uptake data. To determine which region was better, we created logistic regression models for differentiating between all three groups normalizing to brainstem and cerebellum individually and in combination, which resulted in three separate models. To compare the models directly we took only the common regions to all three models and forced them into separate final models. These final models showed that the cerebellum slightly outperformed the brainstem, with c-statistics of 0.826 and 0.823, respectively. Thus, for the remainder of the study, the cerebellum was used as the normalization region.

Partial Volume Correction

ANOVAs for the effects of partial volume correction: To assess the effects of PVC, we examined the change in between-group relationships as determined by Tukey's post-hoc ANOVA tests. Prior to PVC, the majority of regions showed significant changes between normal aging, MCI, and AD (Table 2), with the exception of some cingulate, frontal, visual, and mid-brain regions. The highest degrees of change were observed in the medial temporal lobe, frontal, temporal, and parietal regions. After PVC using residuals, nearly all previously significant regions of hypometabolism were not insignificant, with the exception of the postcentral gyrus. Three regions showed an increase in significant, namely the amygdala, thalamus, and temporal pole. After PVC using the GMM method, there were again mixed effects, with some regions decreasing in significance and other regions increasing in significance. Table 2 outlines the specific details.

Stepwise logistic regression models: To examine the effects of partial volume correction, we created a series of models that differentiated normal aging from MCI, normal aging from AD, MCI from AD, and all three groups from one another before partial volume correction, after correction using a statistically-based residuals approach, and after correction using the GMM. Here we present the results from differentiating normal aging from MCI and normal aging from AD. In all models, age, gender, and education were forced in to control for the variance they may have had on the model. Similar results were found for all four model conditions, thus the other two groups of models for differentiating AD from MCI and all three groups are presented in the supplementary material.

	P-valuenon-PVC	P-value residuals	P-value GMM	MCI vs. normal	AD vs. MCI	AD vs. normal
banks sts*	<.0001	0.6258	0.0013	<, =, =	=, =, =	<, =, <
caudal anterior cingulate	0.6019	0.3254	0.4238	=, =, =	=, =, =	=, =, =
caudal middle frontal	0.7342	0.3182	0.0526	=, =, =	=, =, =	=, =, =
cuneus	<.0001	0.3951	0.2502	<, =, =	=, =, =	<, =, =
entorhinal	0.0007	0.0652	<.0001	=, =, <	=, =, =	<, <, <
frontal pole	<.0001	0.1217	0.2844	<, =, =	<, =, =	<, =, =
fusiform	<.0001	0.1756	0.0276	<, =, <	<, =, =	<, =, =
inferior parietal	<.0001	0.0677	0.0006	<, =, <	<, =, =	<, <, <
inferior temporal	0.0393	0.0074	<.0001	=, =, <	=, <, =	<, <, <
insula	0.0068	0.7174	0.2448	=, =, =	=, =, =	<, =, =
isthmus of the cingulate	0.0879	0.0688	<.0001	=, =, <	=, =, =	=, <, <
lateral occipital	0.1385	0.0516	0.287	=, =, =	=, =, =	=, =, =
lateral orbitofrontal	0.0066	0.3856	0.1115	=, =, =,	=, =, =	<, =, =
lingual	0.0012	0.1966	0.2779	=, =, =,	=, =, =	<, =, =
medial orbitofrontal	<.0001	0.8302	0.1115	=, =, =,	<, =, =	<, =, =
middle temporal	0.0009	0.2141	<.0001	=, =, <	<, =, =	<, =, <
paracentral	0.0001	0.0175	0.016	=, >, =	<, =, >	<, >, >
parahippocampal gyrus	0.0296	0.0078	0.0045	=, =, <	=, =, =	<, <, <
pars opercularis	0.0025	0.3434	0.167	=, =, =	<, =, =	<, =, =
pars orbitalis	0.0016	0.3855	0.253	=, =, =	<, =, =	<, =, =
pars triangularis	<.0001	0.5239	0.2787	<, =, =	<, =, =	<, =, =
pericalcarine	0.8233	0.1407	0.1275	=, =, =	=, =, =	=, =, =
postcentral gyrus	<.0001	0.0399	0.0004	<, =, =	<, =, >	<, =, >
posterior cingulate	0.1326	0.1107	0.0563	=, =, =	=, =, =	=, =, <
precentral gyrus	0.007	0.6059	0.0054	=, =, =	=, =, >	<, =, =
precuneus	0.0345	0.0809	0.0081	=, =, <	=, =, =	<, =, <
rostral anterior cingulate	<.0001	0.3125	0.2364	=, =, =	<, =, =	<, =, =
rostral middle frontal	<.0001	0.2882	0.1706	=, =, =	<, =, =	<, =, =
superior frontal	0.0014	0.5823	0.0581	=, =, =	<, =, =	<, =, =
superior parietal	<.0001	0.7744	0.1286	<, =, =	<, =, =	<, =, =
superior temporal	0.0435	0.3626	0.0282	=, =, <	=, =, =	>, =, =
supramarginal	<.0001	0.2503	0.1262	<, =, =	<, =, =	<, =, =
temporal pole	0.3721	0.0026	0.0097	=, >, <	=, =, =	=, =, =
transverse temporal	<.0001	0.244	0.4129	<, =, =	<, =, =	<, =, =
amygdala	0.4738	<.0001	0.0035	=, <, <	=, <, =	=, <, <
caudate	0.5027	0.4419	0.0717	=, =, =	=, =, =	=, =, =
hippocampus	0.0015	0.162	0.0024	=, =, <	<, =, =	<, =, <
pallidum	0.0147	0.7449	0.0305	=, =, =	=, =, =	<, =, >
putamen	0.1322	0.8875	0.0022	=, =, <	=, =, =	=, =, =
thalamus	0.0403	0.0011	0.1898	=, =, =	=, =, =	<, <, =

*sts = superior temporal sulcus

The first three columns show p-values from ANOVAs comparing uptake in normal aging, MCI, and AD before partial volume correction (PVC) and after PVC using residuals and gray matter mask (GMM)methods. The MCI vs. normal, AD, vs. MCI, and AD vs. normal columns show the specific relationships between groups, whereby '<' indicates a decrease in MCI compared to normal in the first column, AD < MCI, and AD < normal in the three columns, respectively (p > 0.05). Similarly, '>' indicates a significant increase in one group compared to the other, i.e. MCI > normal, AD > MCI, and AD > normal. '=' indicates no significant change in uptake between groups.

Table 2: ANOVA results for FDG uptake changes between normal aging, MCI, and AD subject groups for non-PVC, PVC via cortical thickness + subcortical volume residuals, and PVC via GMM methods.

Models normal aging from MCI: Partial volume correction significantly decreased the ability to predict group membership, regardless of which method was used. The GMM method provided a better model than cortical thickness residuals, both in terms of c-statistic (ROC) and AIC. Prior to partial volume correction, models differentiating between normal aging and MCI provided a c-statistic of 0.810 and AIC of 310 (Table 3). Hosmer-Lemeshow goodness of fit shows that this model is well calibrated (Chi-square = 11.82, p = 0.16). The regions that contributed significant amounts of variance to the model included the entorhinal cortex and the hippocampus. Age was the only other significant predictor. Banks of the superior temporal sulcus, caudal middle frontal, inferior temporal, and superior temporal all contributed variance to the model, but these failed to reach statistical significance.

After correcting using cortical thickness and subcortical volume

MCI versus normal aging	Unit	Odds Ratio (95%CI)	Significance
Non-PVC model			
gender	1	1.855 (0.995 - 3.462)	0.0521
education	2.9586	0.881 (0.655 - 1.185)	0.4027
age	6.5536	0.597 (0.439 - 0.812)	0.001
entorhinal	337.3	0.681 (0.478 - 0.970)	0.0335
banks STS	341.8	0.773 (0.558 - 1.071)	0.1214
caudal middle frontal	876.3	0.742 (0.538 - 1.023)	0.0687
inferior temporal	1279	0.744 (0.520 - 1.064)	0.105
superior temporal	1166.3	1.418 (0.948 - 2.122)	0.0894
hippocampus	532.6	0.345 (0.222 - 0.537)	<0.0001
Cortical thickness + subcortical volume residuals			
gender	1	1.131 (0.645 - 1.985)	0.6677
education	2.9586	0.907 (0.703 - 1.171)	0.4536
age	6.5536	0.815 (0.625 - 1.064)	0.1331
paracentral gyrus	371.1	1.261 (0.949 - 1.674)	0.1094
temporal pole	225.5	1.478 (1.134 - 1.927)	0.0039
amygdala	21.0682	0.667 (0.499 - 0.890)	0.0059
Gray matter mask			
gender	1	1.507 (0.869 - 2.612)	0.1444
education	2.9574	0.857 (0.658 - 1.117)	0.2539
age	6.4782	0.798 (0.611 - 1.041)	0.0965
entorhinal	1349.3	0.549 (0.361 - 0.835)	0.0051
fusiform	1235.9	2.439 (1.342 - 4.434)	0.0035
inferior temporal	1301	0.576 (0.341 - 0.975)	0.04
isthmus of the cingulate	2007.9	0.588 (0.386 - 0.895)	0.0133

The top model shows the non-pvc model (c = 0.810), followed by the cortical thickness + subcortical volume residuals (c = 0.678), and the GMM on the bottom (c = 0.688).

Table 3: Stepwise logistic regression models for differentiating between MCI and normal aging groups before and after partial volume correction.

residuals, the c-statistic was 0.678 and AIC was 361. Hosmer-Lemeshow goodness of fit shows that this model is well calibrated (Chi-square = 7.33, p = 0.50). The difference in c-statistics between non-PVC and residuals-PVC models was statistically significant (p = 0.0008). The regions that were significant predictors were the amygdala and temporal pole. The paracentral gyrus contributed but failed to reach statistical significance.

After correcting using the GMM method, the c-statistic was 0.688 and AIC was 355. Hosmer-Lemeshow goodness of fit shows that this model is well calibrated (Chi-square = 4.29, p = 0.83). The difference in c-statistics between non-PVC and GMM-PVC models was also statistically significant (p = 0.0015). The regions that contributed significantly to the model were the entorhinal, fusiform, inferior temporal, and the isthmus of the cingulate. The c-statistic and AIC both indicate that the GMM method of PVC provides the best model for discriminating MCI from normal aging.

Models differentiating normal aging from AD: Again, partial volume correction significantly decreased the ability to discriminate group and the GMM method provided a better model, both in terms of c-statistic and AIC, than the residuals method of PVC (Table 4). Prior to PVC, differentiating between normal aging and AD subject groups provided a model with a c-statistic of 0.972 and AIC of 110. However, Hosmer-Lemeshow goodness of fit test shows that this model is not well calibrated and may not provide consistent results (Chi-square = 50.86, p < 0.0001). The variables that contributed significantly to the model include precuneus,, rostral anterior cingulate, isthmus of the cingulate, insula, parahippocampal gyrus, hippocampus, and superior parietal. Age was also a significant predictor.

After correcting using cortical thickness and subcortical volume residuals, the c-statistic was 0.864 and AIC was 195. Hosmer-Lemeshow goodness of fit shows that this model is well calibrated (Chi-square = 6.57, p = 0.58). The regions that contributed significantly to the model were the entorhinal, amygdala, parahippocampus, thalamus, paracentral gyrus, and the inferior temporal lobe. Education was a significant predictor, as well.

After correcting using GMM, the c-statistic was 0.879 and AIC was 185. Hosmer-Lemeshow goodness of fit shows that this model is well calibrated (Chi-square = 3.89, p = 0.87). The regions that contributed significantly to the model were the entorhinal, parahippocampus, middle temporal, temporal pole, and the paracentral and post-central gyri. Education was also a significant predictor. Other regions that contributed, although not significantly, to the model were the isthmus of the cingulate and the precuneus.

Direct comparison of models: Because different regions, only some of which are previously attributed to MCI and AD, contributed to each of the models of different correction types, we wanted to verify the efficacy of one correction method over another by allowing only specific brain regions that are consistently affected in MCI and AD into the models. Thus, we re-created models using only the measures of the entorhinal cortex, age, gender, and education in order to directly compare the non-partial volume corrected model and the two PVC models. When comparing MCI and normal aging, c = 0.726 for non-PVC, c = 0.655 for GMM, and c = 0.594 for cortical thickness residuals. When comparing AD and normal aging, c = 0.865 for non-PVC, c = 0.713 for GMM, and c = 0.653 for cortical thickness residuals. Similar results were found for comparing MCI and AD and for differentiating

AD versus normal aging	Unit	Odds Ratio (95%CI)	Significance
Non-PVC model			
gender	1	2.395 (0.664 - 8.639)	0.1822
education	3.2426	0.561 (0.295 - 1.068)	0.0784
age	6.0469	0.203 (0.079 - 0.523)	0.001
frontal pole	131.3	0.621 (0.321 - 1.201)	0.1571
fusiform	1171.9	0.483 (0.229 - 1.019)	0.0561
precuneus	1003.1	0.132 (0.044 - 0.393)	0.0003
rostral anterior cingulate	276.6	2.040 (1.002 - 4.156)	0.0495
isthmus of the cingulate	272.3	2.725 (1.090 - 6.808)	0.0319
insula	660.4	2.997 (1.271 - 7.067)	0.0121
parahippocampal gyrus	305.3	0.441 (0.214 - 0.909)	0.0265
superior parietal	1458.7	3.218 (1.245 - 8.318)	0.0159
hippocampus	621.4	0.012 (0.002 - 0.067)	<0.0001
Cortical thickness + subcortical volume residuals			
gender	1	1.408 (0.575 - 3.45)	0.4541
education	3.2426	0.526 (0.344 - 0.802)	0.0029
age	6.0469	0.703 (0.470 - 1.051)	0.0855
entorhinal	246.9	0.597 (0.391 - 0.912)	0.0171
inferior temporal	1297.9	0.595 (0.368 - 0.961)	0.0339
paracentral gyrus	328.2	2.671 (1.640 - 4.351)	<0.0001
parahippocampal gyrus	183.3	0.505 (0.307 - 0.833)	0.0074
amygdala	23.6388	0.322 (0.195 - 0.530)	<0.0001
thalamus	52.2657	0.612 (0.411 - 0.911)	0.0156
Gray matter mask			
gender	1	1.072 (0.450 - 2.551)	0.8757
education	3.2507	0.574 (0.379 - 0.869)	0.0087
age	5.9331	1.051 (0.705 - 1.567)	0.8079
entorhinal	1508.9	0.103 (0.033 - 0.325)	0.0001
isthmus of the cingulate	2183.6	0.423 (0.120 - 1.495)	0.1818
middle temporal	1686	0.231 (0.106 - 0.501)	0.0002
paracentral gyrus	2075.2	2.471 (1.020 - 5.987)	0.0452
parahippocampal gyrus	1576.2	4.218 (1.220 - 14.579)	0.0229
postcentral gyrus	1575.6	3.685 (1.463 - 9.279)	0.0056
precuneus	2083.1	0.333 (0.102 - 1.087)	0.0684
temporal pole	1264.5	3.130 (1.132 - 8.655)	0.0279

The top model shows the non-pvc model (c = 0.972), followed by the cortical thickness + subcortical volume residuals (c = 0.864), and the GMM on the bottom (c = 0.879).

Table 4: Stepwise logistic regression models for differentiating between AD and normal aging groups before and after partial volume correction.

all three groups. The results of these analyses confirm that GMM provides higher c-statistics than cortical thickness residuals, but lower than non-PVC values.

Discussion

FDG PET can provide vital information in the study of degenerative diseases such as Alzheimer's disease. In the raw, unprocessed form, FDG PET data can be confusing to interpret quantitatively particularly when making comparisons between groups. In order to make these comparisons more straightforward, the uptake in each scan must be normalized to a control region within the scan and corrected for the influence of factors that could adversely impact the data. In this study we first examined which anatomically based ROI could serve as an appropriate region for normalizing FDG uptake across subjects ranging from normal aging, MCI and Alzheimer's disease. These regions were chosen because they are known to show relatively preserved glucose metabolism in AD. Next we looked into a means for addressing partial volume effects, a morphometric feature which can become more pronounced with atrophy in disease states such as Alzheimer's disease. We assessed two methods, one using a gray matter mask to remove the potential confounding effects of the white matter, and the second using a statistical correction based upon using the residuals derived from cortical thickness or subcortical volume depending upon the structure. Specifically we sought to identify the best correction factors for partial volume errors and normalization to obtain the statistical model which best differentiates between FDG uptake in normal aging, MCI, and AD.

Normalization region ANOVA

Our results suggest that the cerebellum is the best region to use for when normalizing scans in normal aging, MCI and early stage AD. Previous studies have found that the cerebellum provides better accuracy than cerebral global mean [5,6,24] which may further improve if the vermis or superior cerebellar regions are targeted [7]. The further subdivision of cerebellum into its component parts is beyond the scope of this study, although it would be interesting to examine in a large sample whether the vermis and superior cerebellar portions

provide better normalization than the cerebellum taken as a whole. Pathologically, the cerebellum, including the cerebellar hemispheres, posterior cerebellum, and vermis, shows an overall decrease in volume corresponding to disease severity in AD [25]. This has further been shown to related to a loss of Purkinje cell density, neuronal loss, and gliosis in the vermis, with relative sparing of the lateral portions of the cerebellar hemispheres [26]. Because the vermis was particularly affected in these two studies, it remains to be seen whether it can consistently be shown to be a better normalization region than the whole cerebellum.

The pre- and post-central gyri have recently been implicated as the best region for normalization [3] in AD. One of the requirements of a normalization region in this study was a lack of significant metabolic changes with disease. Both the pre- and post-central gyri showed significant hypometabolism in the AD group, which is consistent with a previous study [4], suggesting that these primary somatosensory cortices are not appropriate normalization regions when AD subjects are being used, despite relative sparing of pathology. Yakushev published a follow-up study [24] whereby normalization to the cerebral global mean, primary somatosensory cortex, and a cluster-derived region were compared for classification accuracy of normal aging, MCI, and AD groups. The cluster-derived region provided the greatest relative decreases in metabolism in MCI and AD groups, as well as provided the highest classification accuracy for both groups. In this study, a figure indicates that the cluster region was in the cerebellum, providing strong evidence that the cerebellum may indeed outperform the somatosensory cortices for normalization. Neither of the Yakushev studies use partial volume correction, thus the findings may be based on a mixture of metabolic and morphometric changes.

The brainstem has been used previously as a normalization reference region [27]. This is not a typical region, although the pons, a substructure within it, is. We chose to assess the brainstem, rather than the pons for a number of reasons. The pons is not a standard output in standard atlases, such as that found in FreeSurfer, but the brainstem is. Atlases provide a number of advantages for generating ROIs, particularly within large datasets, such as the ADNI. With the increasing number of large-scale studies, it is important to study how these readily available atlas-based ROIs are able to normalize. The expansion of these atlases to include the pons in the standard regions may assist in future studies. One of the difficulties in segmenting the pons from the rest of the brainstem tissue is that in MRI, the posterior border is not well delineated, making is susceptible to increased variability. Pons normalization is also problematic because its signal is noisy [28] making it vulnerable to increasing the variance of normalized regions. Although discreet regions within the brainstem, such as the locus coeruleus, are susceptible to early AD pathology [29] this did not cause overall changes in glucose metabolism within out sample.

Although the thalamus may have shown relative preservation of glucose metabolism in AD, we found that after PVC the thalamus showed significant hypometabolism in the AD group. Similarly, Minoshima et al. [4] found significant disease-related differences in the thalamus; however, this was without any partial volume correction making it difficult to discern whether this was due to changes in metabolism, volume, or both. In this study we only found thalamus differences after correcting for partial volume effects suggesting that this difference is the result of a true reduction in metabolism and not a change in volume.

We did not assess the contribution of a whole brain scaling factor [2,13] because the widespread changes in FDG PET metabolism associated with AD observed violates one of the principles of choosing a reference region for normalization, namely that the region should not show disease-related changes. It has been suggested that the cerebellum provides better normalization for comparisons between normal aging and either MCI or AD, but that cerebral global mean may be better for comparing MCI and AD groups. Because of widespread hypometabolism in both MCI and AD, we did not assess the efficacy of cerebral global mean as a normalization factor. Although we would expect that using the whole brain would have resulted in regions of artificially augmented metabolism, as has been reported previously [30,31].

Taken together, our findings on the pre- and post-central gyrus and thalamus affirm the importance of correcting for partial volume effects. It also underscores the importance of testing the reference region for disease-related effects in the subject sample to ensure its validity as a reference region.

Partial Volume Correction

Many studies that have looked at choosing a normalization region have not corrected for partial volume effects and those which have looked at PVC have not examined which brain region was best for normalization in their sample. Since these two factors can interact, it was important to look at both factors simultaneously in a sample that included not only normal aging and AD subjects, but also MCI subjects, which have often been neglected in these types of study.

We found that the adapted GMM method, which corrects the data at the image level, provided a higher index of discriminability in all group comparisons than using a statistical correction with cortical thickness or subcortical volume residuals. In addition, the regions that were chosen by the analysis for inclusion into the final models with the GMM method also are more consistent with those typically impacted by Alzheimer's disease.

The changes in Tukey's post-hoc analysis indicate that there is an interaction between partial volume correction and FDG metabolism amongst the three groups. The decrease in significance observed throughout the cortex would indicate that the FDG data prior to correcting for partial volume was likely contaminated by atrophy, particularly in the AD group, as reflected by the low uptake values. We would like to point out that in a number of instances, regions typically associated with AD, (i.e. the hippocampus) were in the non-PVC model, but not in the PVC model. This does not necessarily imply that the regions no longer showed significant hypometabolism with disease. The logistic models indicate which variables contribute unique amounts of variance to the discrimination of two or more subject groups. Thus, there can still be significant disease-associated hypometabolism in a region that is not in the logistic model. Thus, the hippocampus may contribute variance to differentiating normal aging from MCI, but if the variance overlaps with a region that contributes more variance, such as the entorhinal cortex, then the hippocampus would not be included in the model, but the entorhinal cortex would be. This study is the first of its kind to our knowledge to identify both the optimal choice of reference region for normalization and partial volume correction method in a large dataset from multiple research centers. Nonetheless there are a few limitations to our study. This is a cross sectional study, which may limit the application to longitudinal studies. The ADNI participants are also largely recruited from memory centers, which may limit the applicability of the results to primary care facilities. In addition, the brainstem was used as a surrogate for the pons. Certain regions

within the brainstem (e.g. locus coeruleus) show early accumulation of AD pathology [29]. Although disease-related changes have been observed in portions of the brainstem, our ANOVA results indicate that there were no significant influences of AD pathology on the FDG metabolism in the brainstem. Despite the shortcomings of this study, the results confirm that normalization to the cerebellum provides the best discrimination between groups and that partial volume correction using a gray matter mask performs better than using cortical thickness residuals. Because the sample is so large and spans multiple research centers and scanner platforms, the results are likely to translate to smaller studies well.

Conclusion

We sought to determine which choice of reference region for normalization provided the best model for discriminating between normal aging, MCI, and AD subject groups and also to determine the effects of partial volume correction on the statistical models. Partial volume correction is necessary for identifying which brain regions show true changes in FDG uptake with disease progression independent of any changes in MRI morphometry. Our results indicate that for this study sample, the cerebellum was the best region for normalization. The best models were those in which there was no partial volume correction, but by correcting for partial volume effects we can be certain that the FDG changes are from decreased metabolism and not influenced by atrophy. Out of the two PVC methods tested (cortical thickness and subcortical volume residuals and GMM), we found that the GMM provides a higher index of discriminability, as measured with the c-statistic. Not only that, but the regions that were predictors in the GMM model better concurred with the literature on which regions are involved at the MCI and early AD disease stages. In addition, we can conclude that partial volume correction with the GMM diminished the age-related changes observed in the normal aging population. These results together suggest that perhaps the GMM is better than using residuals for FDG PET data processed in Freesurfer.

References

1. Desgranges B, Baron JC, Lalevée C, Giffard B, Viader F, et al. (2002) The neural substrates of episodic memory impairment in Alzheimer's disease as revealed by FDG-PET: relationship to degree of deterioration. Brain 125: 1116-1124.

2. Dukart J, Mueller K, Horstmann A, Vogt B, Frisch S, et al. (2010) Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. Neuroimage 49: 1490-1495.

3. Yakushev I, Landvogt C, Buchholz HG, Fellgiebel A, Hammers A, et al. (2008) Choice of reference area in studies of Alzheimer's disease using positron emission tomography with fluorodeoxyglucose-F18. Psychiatry Res 164: 143-153.

4. Minoshima S, Frey KA, Foster NL, Kuhl DE (1995) Preserved pontine glucose metabolism in Alzheimer disease: a reference region for functional brain image (PET) analysis. J Comput Assist Tomogr 19: 541-547.

5. Guenther T, Küntzelmann A, Haberkorn U, Essig M, Schröder J, Schönknecht P (2011) Comparison between global and cerebellar intensity normalization of FDG PET images in patients with mild cognitive impairment or Alzheimer's disease. Pharmacopsychiatry 44: 21-A45.

6. Küntzelmann A, Guenther T, Haberkorn U, Essig M, Giesel F, et al. (2013) Impaired cerebral glucose metabolism in prodromal Alzheimer's disease differs by regional intensity normalization. Neurosci Lett 534: 12-17.

7. Rasmussen JM, Lakatos A, van Erp TG, Kruggel F, Keator DB, et al. (2012) Empirical derivation of the reference region for computing diagnostic sensitive Â¹â‚fluorodeoxyglucose ratios in Alzheimer's disease based on the ADNI sample. Biochim Biophys Acta 1822: 457-466.

8. Rousset OG and Zaidi H (2006) Correction for Partial Volume Effects in Emission Tomography. In Quantitative Analysis in Nuclear Medicine Imaging, H. Zaidi, ed. Kluwer Academic Publishers, Boston.

9. Meltzer CC, Kinahan PE, Greer PJ, Nichols TE, Comtat C, et al. (1999) Comparative evaluation of MR-based partial-volume correction schemes for PET. J Nucl Med 40: 2053-2065.

10. Meltzer CC, Zubieta JK, Brandt J, Tune LE, Mayberg HS, et al. (1996) Regional hypometabolism in Alzheimer's disease as measured by positron emission tomography after correction for effects of partial volume averaging. Neurology 47: 454-461.

11. Park HJ, Lee JD, Chun JW, Seok JH, Yun M, et al. (2006) Cortical surface-based analysis of 18F-FDG PET: measured metabolic abnormalities in schizophrenia are affected by cortical structural abnormalities. Neuroimage 31: 1434-1444.

12. Rousset OG, Ma Y, Evans AC (1998) Correction for partial volume effects in PET: principle and validation. J Nucl Med 39: 904-911.

13. Samuraki M, Matsunari I, Chen WP, Yajima K, Yanase D, et al. (2007) Partial volume effect-corrected FDG PET and grey matter volume loss in patients with mild Alzheimer's disease. Eur J Nucl Med Mol Imaging 34: 1658-1669.

14. Yanase D, Matsunari I, Yajima K, Chen W, Fujikawa A, et al. (2005) Brain FDG PET study of normal aging in Japanese: effect of atrophy correction. Eur J Nucl Med Mol Imaging 32: 794-805.

15. Walhovd KB, Fjell AM, Dale AM, McEvoy LK, Brewer J, et al. (2010) Multi-modal imaging predicts memory performance in normal aging and cognitive decline. Neurobiol Aging 31: 1107-1121.

16. Ishii K (2002) Clinical application of positron emission tomography for diagnosis of dementia. Ann Nucl Med 16: 515-525.

17. Matsuda H, Ohnishi T, Asada T, Li ZJ, Kanetaka H, et al. (2003) Correction for partial-volume effects on brain perfusion SPECT in healthy men. J Nucl Med 44: 1243-1252.

18. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12: 189-198.

19. Morris JC (1993) The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43: 2412-2414.

20. Auer S, Reisberg B (1997) The GDS/FAST staging system. Int Psychogeriatr 9 Suppl 1: 167-171.

21. Reisberg B, Ferris SH, de Leon MJ, Crook T (1988) Global Deterioration Scale (GDS). Psychopharmacol Bull 24: 661-663.

22. Wechsler D (1987) WMS-R: Wechsler Memory Scale--Revised manual Psychological Corp, Harcourt Brace Jovanovich.

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

24. Yakushev I, Hammers A, Fellgiebel A, Schmidtmann I, Scheurich A, et al. (2009) SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. Neuroimage 44: 43-50.

25. Baldaçara L, Borgio JG, Moraes WA, Lacerda AL, Montaño MB, et al. (2011) Cerebellar volume in patients with dementia. Rev Bras Psiquiatr 33: 122-129.

26. Sjöbeck M, Englund E (2001) Alzheimer's disease and the cerebellum: a morphologic study on neuronal and glial changes. Dement Geriatr Cogn Disord 12: 211-218.

27. Walhovd KB, Fjell AM, Amlien I, Grambaite R, Stenset V, et al. (2009) Multimodal imaging in mild cognitive impairment: Metabolism, morphometry and diffusion of the temporal-parietal memory network. Neuroimage 45: 215-223.

28. Fletcher PT, Kumar A, Wang AY, Jagust WJ, Chen K, Reiman EM, Weiner MW,

Foster NL, Alzheimer's Disease Neuroimaging Initiative (2010) Regression-to-pons normalization of FDG-PET improves discrimination of Alzheimer's disease from healthy aging. Alzheimers Dement. J. Alzheimers Assoc. : S11–S12.

29. Braak H, Del Tredici K (2011) Alzheimer's pathogenesis: is there neuron-to-neuron propagation? Acta Neuropathol 121: 589-595.

30. Borghammer P, Cumming P, Aanerud J, Gjedde A (2009) Artefactual subcortical hyperperfusion in PET studies normalized to global mean: lessons from Parkinson's disease. Neuroimage 45: 249-257.

31. Borghammer P, Jonsdottir KY, Cumming P, Ostergaard K, Vang K, et al. (2008) Normalization in PET group comparison studies--the importance of a valid reference region. Neuroimage 40: 529-540.

This article was originally published in a special issue, **Neurodegenerative Disorders** handled by Editor(s). Dr. Aleksandar Videnovic, Northwestern University, USA