NeuroImage 56 (2010) 782–787

Contents lists available at ScienceDirect

NeuroImage

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

Robustness of correlations between PCA of FDG-PET scans and biological variables in healthy and demented subjects

P.J. Markiewicz [a,*], J.C. Matthews [b], J. Declerck [c], K. Herholz [b,d]

[a] School of Cancer and Imaging Sciences at the University of Manchester, Wolfson, Molecular Imaging Centre, Manchester, England
[b] Research School of Translational Medicine at the University of Manchester, Wolfson, Molecular Imaging Centre, Manchester, England
[c] Siemens Molecular Imaging, Oxford, England
[d] Department of Neurology, University of Cologne, Germany

ARTICLE INFO

Article history:
Received 1 December 2009
Revised 18 May 2010
Accepted 25 May 2010
Available online 2 June 2010

Keywords:
Correlation
PCA
Robustness
Significance
PET
Alzheimer's disease

ABSTRACT

In neuroimaging it is helpful and useful to obtain robust and accurate estimates of relationships between the image derived data and separately derived covariates such as clinical and demographic measures. Due to the high dimensionality of brain images, complex image analysis is typically used to extract certain image features, which may or may not relate to the covariates. These correlations which explain variance within the image data are frequently of interest.

Principal component analysis (PCA) is used to extract image features from a sample of 42 FDG PET brain images (19 normal controls (NCs), 23 Alzheimer's disease (AD) patients). For the first three most robust PCs, the correlation of the PC scores with: i) the Mini Mental Status Exam (MMSE) score and ii) age is examined. The key aspects of this work is the assessment of: i) the robustness and significance of the correlations using bootstrap resampling; ii) the influence of the PCA on the robustness of the correlations; iii) the impact of two intensity normalization methods (global and cerebellum).

Results show that: i) Pearson's statistics can lead to overoptimistic results. ii) The robustness of the correlations deteriorate with the number of PCs. iii) The correlations are hugely influenced by the method of intensity normalization: the correlation of cognitive impairment with PC1 are stronger and more significant for global normalization; whereas the correlations with age were strongest and more robust with PC2 and cerebellar normalization.

Introduction

Functional neuroimaging can provide very useful insight into brain function which significantly relates to cognitive skills and behaviour. In addition to imaging, in neuroscience there are many different non-imaging variables of clinical nature (e.g., the MMSE (Mini Mental State Exam)) and demographical nature (e.g., age).

In Alzheimer's disease (AD) the cognitive function as well as memory are impaired resulting in lower scores of MMSE. ^{18}F-2-Fluoro-2-Deoxy-D-Glucose (FDG) PET allows measurement of regional cerebral glucose metabolism whose reductions correlate closely with impairment of those cognitive functions that are supported by these brain regions especially in the parietotemporal association cortex and also in the frontal lobe (Haxby et al., 1990; Herholz, 2006).

It is very beneficial to find accurate estimates of possible relationships between image data and the clinical and/or demographic covariates (Siedlecki et al., 2009; Salmon et al., 2007). In order to find robust relationships, a number of aspects relating to the nature of the available data has to be considered: i) Sample sizes are usually very small thus reducing the robustness of the statistical analysis. ii) Measurements are always accompanied by noise. iii) It is possible to adapt to some image features not representative of the population. iv) High image dimensionality—the number of voxels exceeds the sample size.

Therefore, multivariate techniques such as principal component analysis (PCA) are often applied to extract useful information and reduce the image dimensionality (Zuendorf et al., 2003; Habeck et al., 2008; Salmon et al., 2007; Markiewicz et al., 2009; Eidelberg, 2007; Scarmeas et al., 2004).

This work concerns the robustness of relationships (correlations) between the PC scores and covariates such as MMSE and age. Estimates of such correlations are subject to sampling randomness to which the PCA itself is also subject. Therefore, to account for the different sources of randomness, the sampling distributions of the correlations are estimated using bootstrap resampling (Efron and Tibshirani, 1993). The significance of the correlations are found using bootstrap resampling and traditional Pearson's statistic.

A very important effect on the correlation analysis is the intensity normalization of the PET images. Intensity normalization ensures that

* Corresponding author.
E-mail address: p.markiewicz@manchester.ac.uk (P.J. Markiewicz).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.05.066

sources of undesired variability do not have impact on the image analysis (Yakushev et al., 2009; Arndt et al., 1996). Two intensity normalizations are investigated: the cerebral global mean and cerebellar normalization.

Subjects, Methods and Theory

The sample consists of $N=42$ subjects from one site (Cologne, Germany), 19 normal controls (NC) and 23 AD patients. The image preprocessing pathway is shown in Fig. 1. More details can be found in Markiewicz et al. (2009).

Intensity normalization

Similarly to Arndt et al. (1996) each voxel x_i with a given intensity is normalized (divided) by the mean value of intensities of voxels from a given unaffected region:

$$x_i^{(\mathrm{norm})} = \frac{x_i}{\mathrm{mean}\left(x_{i\in\{\mathrm{region}\}}\right)}, \tag{1}$$

where the {region} corresponds to the chosen unaffected region. In this work the regions used are: i) the whole brain resulting in global normalization and ii) cerebellum.

Although these two methods are represented by the same general equation, they have different effects on the normalized data which is well illustrated in Fig. 2. Hence, it is expected that the normalization to the cerebellum will have different impact on the correlations, their robustness and significance in comparison to the global normalization.

Multivariate image analysis (PCA)

Principal components (PCs) are the eigenvectors of the covariance matrix of the data matrix (**X**) consisting of column vectors representing spatially and intensity normalized images. PCs are found using singular value decomposition (SVD) or non-linear iterative partial least squares (NIPALS) (Markiewicz et al., 2009).

The number of retained PCs is chosen based on previous findings for the given sample (Markiewicz et al., 2009) indicating that only the first three or four PCs can be treated as robust (here the first three PCs are used).

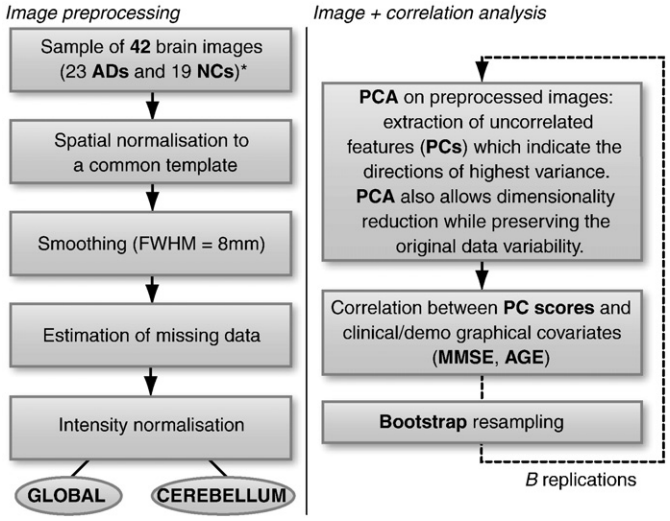

Fig. 1. Image preprocessing (left) and image analysis (right) pathways. Image analysis and estimation of correlation between PC scores and MMSE and age are repeated B times using bootstrap resampling.

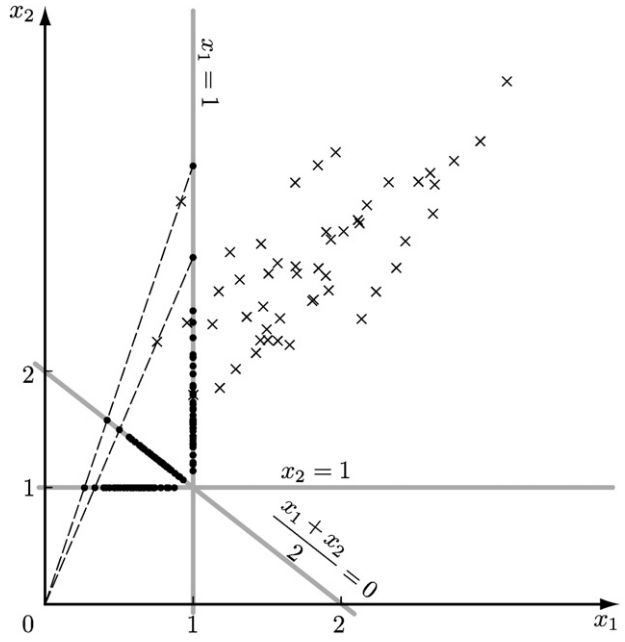

Fig. 2. Three normalization methods illustrated on a two-voxel image (voxels x_1 and x_2). Normalizing by voxel x_1 (which would correspond to a chosen region) results in projecting all voxel values onto line (one-dimensional hyperplane) $x_1=1$; likewise normalizing by voxel x_2 results in projecting all points onto line $x_2=1$. If global mean normalization is used all points are projected onto $(x_1+x_2)/2=0$. The projection is illustrated by a dashed lines for the two extreme points on the left in the plot. Note the different dispersion and variability for the three normalizations shown, e.g., for the x_1 normalization two cases could be recognized as outliers whereas they could not be so viewed for the two other normalizations. Note also that all the normalized points, regardless of the normalization used, lie on lines intersecting the origin of the coordinate system (see the dashed lines).

PC scores which are used in the correlation analysis are found by projecting the original image vectors onto the new PCA subspace spanned by the PCs. That is, matrix of PC scores **S** is found by projecting the image data matrix **X** onto the set of PCs represented by matrix **V**, i.e,

$$\mathbf{S} = \mathbf{V}^T\mathbf{X}. \tag{2}$$

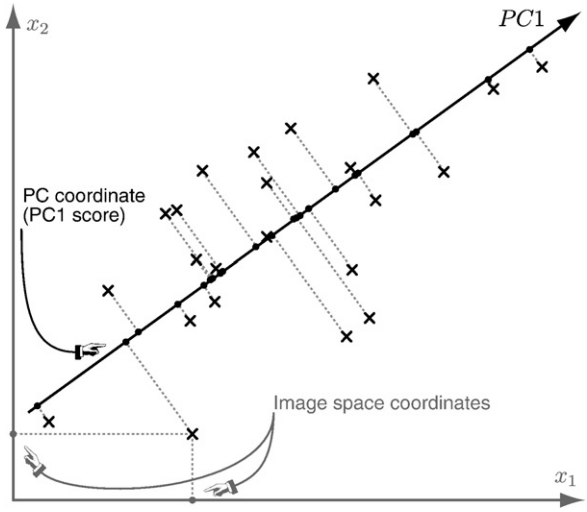

Fig. 3. Illustration of obtaining PC1 scores through projection of the image data points from the original two-dimensional image space onto PC1 which points the direction of the highest variance in the data.

The projection is illustrated in Fig. 3 using image space spanned by just two voxels (x_1 and x_2). The image space data (crosses) are projected onto PC1 resulting in a set of new coordinates—PC1 scores (dots). Such PC scores are used in the correlation analysis.

Reordering and sign changing of the resampled PCs

Since the PCA is to be repeated many times on random bootstrap samples it is possible that the PCA algorithm changes the sign and order of the PCs relative to the original PCs found on the whole sample. Therefore, PCs of every bootstrap sample are reordered and sign changed with reference to the PCs found on the whole sample without resampling. In this way artificial unrobustness in the estimates of the robustness of the correlations can be avoided.

The reordering and sign changing performed here is similar to that of McIntosh and Lobaugh (2004) which is based on Milan and Whittaker (1995).

Correlation analysis

The correlation coefficient r_j for j-th PC score is estimated between the PC score and age or MMSE across subjects according to

$$r_j = \frac{\sum_{i=1}^{N_g}\left(P_{ij}-\bar{P}_j\right)(C_i-\bar{C})}{\left[\sum_{i=1}^{N_g}\left(P_{ij}-\bar{P}_j\right)^2\sum_{i=1}^{N_g}(C_i-\bar{C})^2\right]^{1/2}}, \quad (3)$$

where P_{ij} is the j-th PC score for i-th subject, C_i represents age or MMSE for the i-th subject, P_j is the mean of the j-th PC score across subjects, C is the mean of age or MMSE and N_g is the number of the subject either in the AD group ($g=1$) or NC group ($g=2$).

Significance of the correlations based on parametric statistic

To test whether the correlation r_j is significantly different from zero the parametric test statistic is used,

$$T_j = \frac{r_j\sqrt{n-2}}{\sqrt{1-r_j^2}}, \quad (4)$$

which has the t distribution with n-2 degrees of freedom under the null hypothesis that $r_j=0$.

Robustness and significance of correlations using bootstrap resampling

Each bootstrap sample is formed by uniformly drawing N (the number of subjects in the original sample) instances from the original sample with stratification to reflect the presence of the two populations, AD and NC (Efron and Tibshirani, 1993).

Then for each bootstrap sample the correlation coefficient is found forming a sampling distribution consisting of $B=2000$ replications. Apart from assessing the robustness, the estimated distributions are used to derive the probabilities (p-values) that the correlations came from distributions whose mean is zero. Since the distributions are affected not only by the sample itself but also by the randomness of the PCs, the assessment of the robustness and significance have been decomposed into three cases:

1. PC only. Only PCs are determined anew for every bootstrap sample. The correlations are found between age/MMSE and PC scores for the whole sample without resampling. The influence of the PCs on the correlations is assessed.
2. Correlation only. PCs are found based on the whole sample without resampling and kept constant. The correlations are found between PC scores and age/MMSE of the subjects for every bootstrap sample. The influence of the sample on the correlations is assessed.
3. PC + correlation. Both PCs and correlations are varying in each bootstrap sample. The two above points are combined to result in the most realistic resampling.

The above can be better understood when considering Fig. 3. The randomness of PC1 scores can result from changing the direction of the PC1 (case 1); or from having a different random sample (represented by different points—crosses) projected on the PC1 when it's constant (case 2); or when the two are randomly changing during resampling (case 3).

The correlations between PC scores and age are decomposed into the NC and AD group and investigated separately. The correlations with MMSE are only considered for the AD group.

Results and discussion

The distribution of the correlations for the first three PCs, the three cases of resampling and the two intensity normalization is presented in Fig. 4. The significance levels presented in terms of p-values are shown in bar plots in Fig. 5 for the global normalization and in Fig. 6 for the cerebellar intensity normalization. The significance levels are shown for the three resampling cases (white to dark grey bars) and compared to the traditional t-statistic significance testing (black bars). Note the clear links between the distributions in Fig. 4 and the p-values in Figs. 5 and 6, i.e., the p-values can be directly extracted from the distributions—the further away are the distributions from the zero level the lower the p-value resulting in more significant correlations.

In the bottom of Figs. 5 and 6 are shown the transverse, coronal and sagittal images of the first three PCs for the two intensity normalizations. Note that the global glucose metabolism reduction in AD is visible in PC1 and PC2 with the cerebellar normalization (not visible with global normalization).

The usefulness of robustness on top of statistical significance can be seen in Fig. 4. For example, in the middle row (age in NCs) and middle column (PC2) and for cerebellar normalization (light grey) the distributions clearly indicate correlations significantly different from zero. However, it can be seen that case 'PC' is the most robust (the narrowest distribution) and the 'PC + CORR' case is the least robust of the three. Therefore, robustness provides a complementary information about the distributions significantly different from zero.

From Fig. 4 can be noticed that for the first PC (PC1), both normalizations and all groups 'PC' resampling produces the most robust distributions whereas 'PC + CORR' resampling produces the least robust distributions. This indicates that PC1 is relatively robust and that the robustness of the correlations is more affected by the resampling of the correlations ('CORR' cases). However, for further PCs, especially PC3, the randomness introduced by the PCA becomes dominant (see PC3 for age in NCs).

Similarly, based on Figs. 5 and 6, the significance of the correlations tends to be greater for 'PC' resampling compared to 'PC + CORR' resampling. Insignificant correlations are found to be also unrobust. Importantly, the resampling 'PC + CORR', which estimates the reality, tend to yield less significant correlations (especially true for global normalization). This is due to the fact that the distributions become wider as the two sources of randomness ('PC' and 'CORR') are included in the resampling procedure (cf. Fig. 4).

Further, an important thing can be noticed from the presented results: the standard t-statistic for the Pearson's correlation coefficient can fail in the cases where there is more randomness coming from other sources like the PCA. It is particularly manifest for PC2 and PC3 with the global normalization (see Fig. 5) and also with the cerebellar normalization for PC2 [only in MMSE(AD) case] and PC3. The reason for the overoptimistic result with the standard t-statistic is that it accounts for the randomness of the sample but does not account for

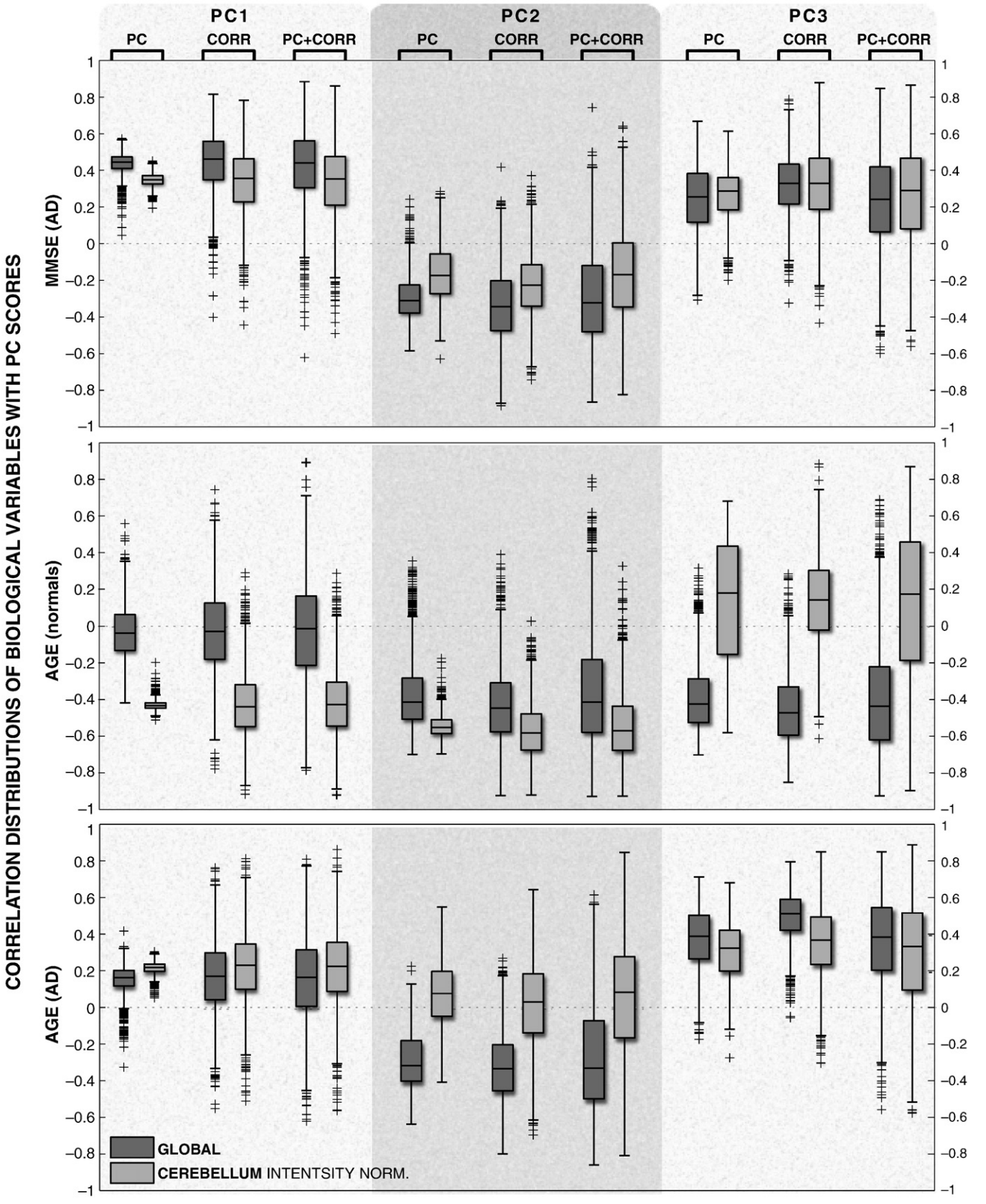

Fig. 4. Box plots of the distributions of the correlation coefficients between biological variables and PC1 (left column), PC2 (middle column) and PC3 (right column). In each column the three cases have been shown: PC, correlation (CORR) and PC + correlation(PC + CORR) resampling. In each of the cases two intensity normalization methods are shown: global (darker grey) and cerebellar (lighter grey) normalization. The correlations are found for MMSE in AD subjects (top row), AGE in NC subjects (middle row) and AGE in AD subjects (bottom row).

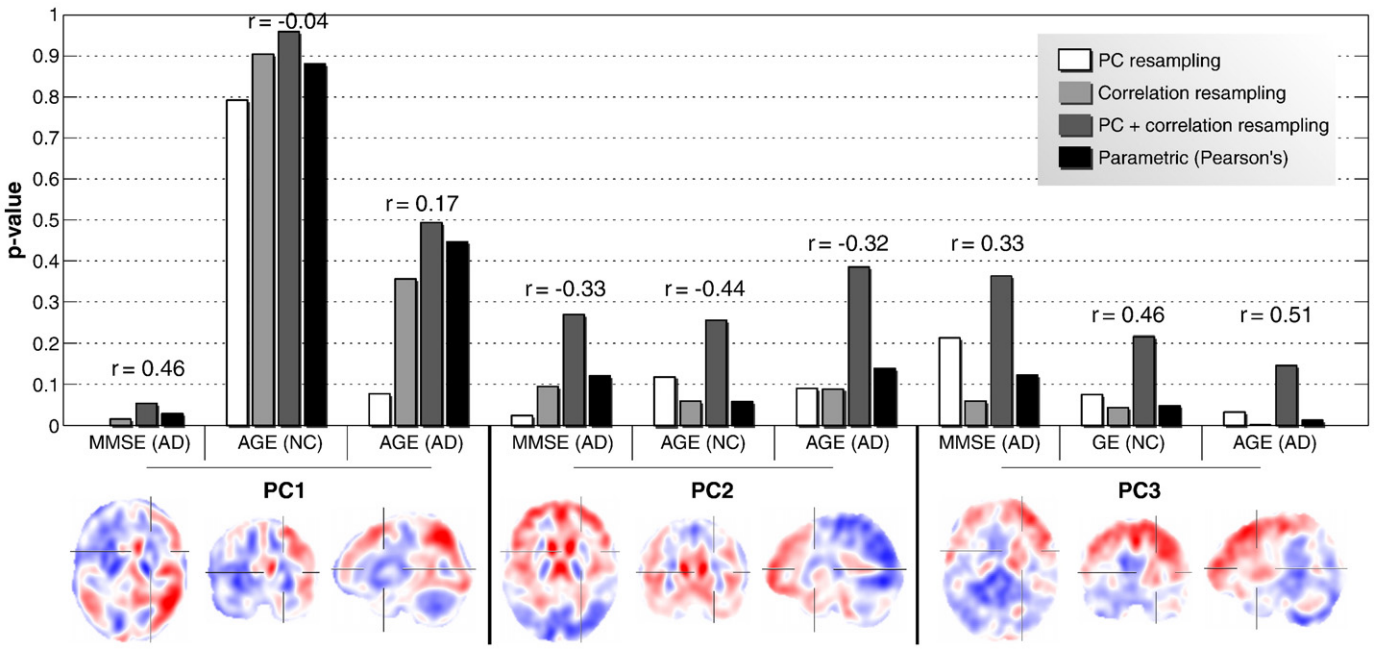

Fig. 5. Comparison of p-values derived from bootstrap distributions for the three cases (PC, correlation and PC + correlation resampling) and the standard t-statistic for the global intensity normalization. The correlation coefficients for the whole dataset without resampling are shown on top of the bar groups.

the complex image analysis (PCA) involved in the correlations. When the PCA is more stable in the case of PC1 the *t*-statistic is quite close to the bootstrap. But when the PCA becomes less stable, as it is for PC2 and PC3, it fails to account for the randomness of the PCs and gives overoptimistic results.

It can also be noted that PC1 is robustly correlated with MMSE reflecting the main contrast between AD patients and NCs. Higher and more significant correlations with MMSE are achieved for global normalization. Correlations with age are higher and more robust with cerebellar normalization for PC1 and PC2 in the NC group, whereas in the AD group the correlations with age are weak and unrobust.

Conclusion

The significance and robustness of the correlations can be assessed with resampling techniques. For the less robust PCs (beginning with the third PC for the given sample) correlations assessed using standard Pearson's statistic (p-values) can be overoptimistic in cases where complex image analysis is involved. This is so due to the fact that the sample size is limited; however, it can be expected that for considerably bigger sample sizes satisfying robustness would be observed with greater number of PCs. Further, the above result has shown that the correlations between PC scores and age as well as

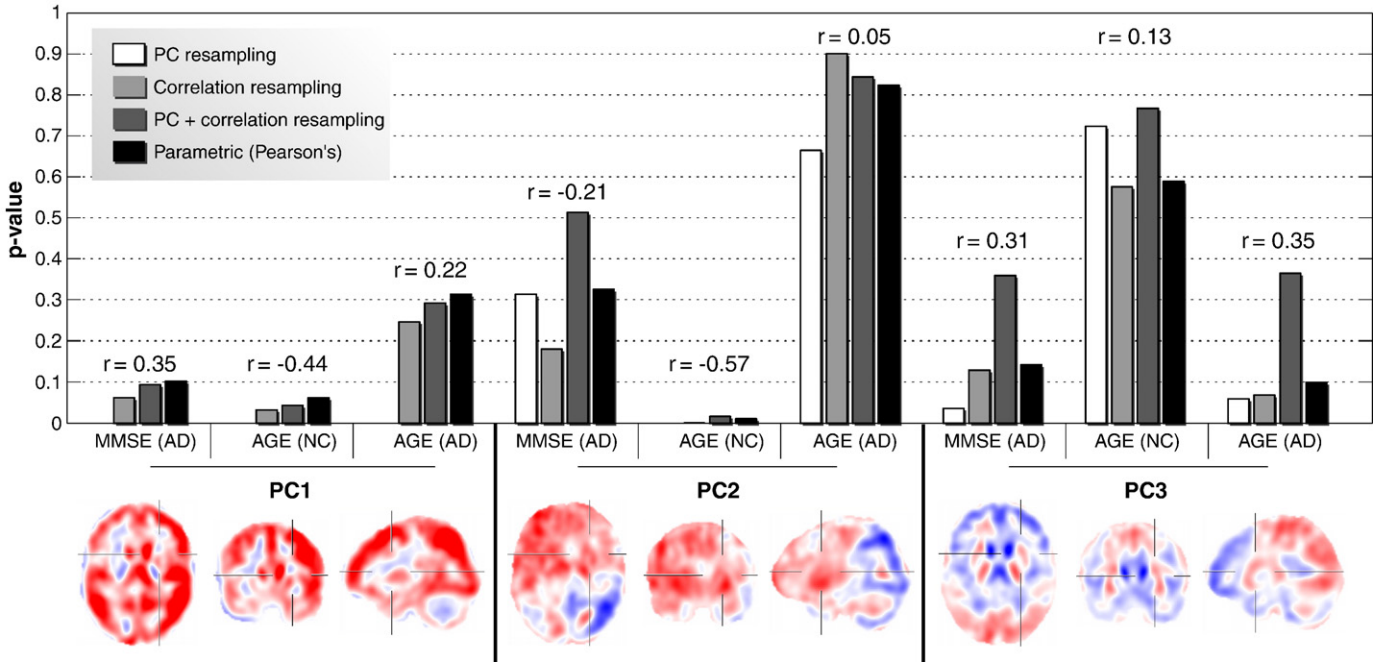

Fig. 6. Comparison of p-values derived from bootstrap distributions for the three cases (PC, correlation and PC + correlation resampling) and the standard t-statistic for the cerebellar intensity normalization. The correlation coefficients for the whole dataset without resampling are shown on top of the bar groups.

MMSE are hugely influenced by the method of intensity normalization and are different for NC and AD subjects. The correlation of PC1 and cognitive impairment (reflected in MMSE) showed stronger, more robust and significant association for global normalization, whereas correlations with age were strongest and more robust with PC2 and cerebellar normalization.

Acknowledgments

This work is funded by Siemens Molecular Imaging and the European Network of Excellence DiMI (LSHBCT-2005-512146)]. The FDG PET data comes from a European multicenter study NEST-DD (CLRT-1999-02178).

References

Arndt, S., Cizadlo, T., O'Leary, D., Gold, S., Andreasen, N.C., 1996. Normalizing counts and cerebral blood flow intensity in functional imaging studies of the human brain. Neuroimage 3, 175–184.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall/CRC.

Eidelberg, D., 2007. The assessment of neurological systems with functional imaging. Brain Lang. 102, 192–199.

Habeck, C., Foster, N.L., Perneczky, R., Kurz, A., Alexopoulos, P., Koeppe, R.A., Drzezga, A., Stern, Y., 2008. Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. Neuroimage 40, 1503–1515.

Haxby, J.V., Grady, C.L., Koss, E., Horwitz, B., Heston, L., Schapiro, M., Friedland, R.P., Rapoport, S.I., 1990. Longitudinal study of cerebral metabolic asymmetries and associated neuropsychological patterns in early dementia of the Alzheimer type. Arch. Neurol. 47, 753–760.

Herholz, K., 2006. FDG PET: Imaging cerebral glucose metabolism with positron emission tomography. In: Herholz, K., Perani, D., Morris, C. (Eds.), The Dementias: Early Diagnosis and Evaluation. Taylor & Francis Group, New York, pp. 229–251.

Markiewicz, P.J., Matthews, J.C., Declerck, J., Herholz, K., 2009. Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease. Neuroimage 46, 472–485.

McIntosh, A.R., Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23, S250–S263.

Milan, L., Whittaker, J., 1995. Application of the parametric bootstrap to models that incorporate a singular value decomposition. Appl. Stat. 44, 31–49.

Salmon, E., Kerrouche, N., Perani, D., Lekeu, F., Holthoff, V., Beuthien-Baumann, B., Sorbi, S., Lemaire, C., Collette, F., Herholz, K., 2007. On the multivariate nature of brain metabolic impairment in Alzheimer's disease. Neurobiol. Aging 30, 186–197.

Scarmeas, N., Habeck, C.G., Zarahn, E., Anderson, K.E., Park, A., Hilton, J., Pelton, G.H., Tabert, M.H., Honig, L.S., Moeller, J.R., Devanand, D.P., Stern, Y., 2004. Covariance PET patterns in early Alzheimer's disease and subjects with cognitive impairment but no dementia: utility in group discrimination and correlations with functional performance. Neuroimage 23, 35–45.

Siedlecki, K.L., Habeck, C.G., Brickman, A.M., Gazes, Y., Stern, Y., 2009. Examining the multifactorial nature of cognitive aging with covariance analysis of postiron emission tomography data. J. Int. Neuropsychol. Soc. 15, 973–981.

Yakushev, I., Hammers, A., Fellgiebel, A., Schmidtmann, I., Scheurich, A., Buchholz, H.-G., Peters, J., Bartenstein, P., Lieb, K., Schreckenberger, M., 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.

Zuendorf, G., Kerrouche, N., Herholz, K., Baron, J.-C., 2003. Efficient principal component analysis for multivariate 3D voxel-based mapping of brain functional Imaging Data Sets as Applied to FDG-PET and Normal Aging. Hum. Brain Mapp. 18, 13–21.