Z-score maps from low-dose 18F-FDG PET of the brain in neurodegenerative dementia

Fällmar, David; Lilja, Johan; Danfors, Torsten; Kilander, Lena; Iyer, Victor; Lubberink, Mark; Larsson, Elna-Marie; Sörensen, Jens

Published in:
American Journal of Nuclear Medicine and Molecular Imaging

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Am J Nucl Med Mol Imaging 2018;8(4):239-246
www.ajnmmi.us /ISSN:2160-8407/ajnmmi0078334

Original Article

Z-score maps from low-dose ^{18}F-FDG PET of the brain in neurodegenerative dementia

David Fällmar[1], Johan Lilja[2,3], Torsten Danfors[2], Lena Kilander[4], Victor Iyer[5], Mark Lubberink[2], Elna-Marie Larsson[1*], Jens Sörensen[2*]

[1]Department of Surgical Sciences, Radiology, Uppsala University, Uppsala, Sweden; [2]Department of Surgical Sciences, Nuclear Medicine and PET, Uppsala University, Uppsala, Sweden; [3]Hermes Medical Solutions, Stockholm, Sweden; [4]Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden; [5]Department of Nuclear Medicine, Aalborg University Hospital, Aalborg, Denmark. [*]Equal contributors.

Received April 22, 2018; Accepted June 14, 2018; Epub August 20, 2018; Published August 30, 2018

Abstract: Neuroimaging is a central part of diagnostic work-up of patients with suspected neurodegenerative disease. FDG-PET can reveal pathological changes earlier and more reliably than morphological imaging. Diagnostic accuracy can be improved by constructing 3D SSP Z-score maps, showing patterns of significant deficits. During FDG-PET, the subject receives a moderate but not insignificant dose of ionizing radiation, and a dose reduction with retained image quality is desirable. With lower dose, repeated examinations can become a useful tool for monitoring disease progress and potential effects of disease-modifying interventions. The aim of this study was to evaluate Z-maps created from low-dose and normal-dose FDG-PET of the brain, with quantitative and qualitative methods. Nine patients with neurodegenerative disorders were prospectively enrolled and nine age-matched controls were recruited through advertising. All subjects (n=18) underwent two FDG-PET scans on separate occasions; a routine and a low-dose scan. The routine dosage of FDG was 3 MBq/kg, and low dosage was 0.75 MBq/kg. 3D-SSP images showing Z-scores of < -1.96 were created from 10-minute summations. The study was comprised of a quantitative part comparing the Z-scores, and a qualitative part where experienced nuclear medicine specialists visually assessed the images. Regarding the quantitative part, Bland-Altman analysis showed a slight constant bias (0.206). Regarding qualitative discrimination between patients and controls, the performance between normal- and low-dose were equal, both showing 72% sensitivity, 83% specificity and 78% accuracy. In this study, visual assessment of 3D-SSP Z-score maps from low-dose FDG-PET provided diagnostic information highly comparable to normal-dose, with minor quantitative discrepancies.

Keywords: PET, FDG, neuroimaging, neurodegeneration, methodology, radiation

Introduction

In 2015, there were 46.8 million patients with dementia disorders worldwide and according to current prognoses there will be 131.5 million by 2050 [1]. If or when disease-modifying drugs become available, establishing an early diagnosis will be of paramount importance. During recent years, neuroimaging has become a central part in the diagnostic work-up of patients with suspected neurodegenerative disease, to exclude treatable causes and to support clinical findings [2]. Morphological changes such as atrophy are evident late in the disease progress, and physiological markers stand a better chance for establishing an early diagnosis [3].

PET examination using ^{18}F-Fluorodeoxyglucose (FDG) reflects transport of glucose into cells, and energy-demanding synaptic activity leads to increased regional glucose uptake. Tracer uptake is mainly driven by basal neuronal activity [3], and regions of reduced activity are visualized as areas of "hypometabolism". Several overviews on typical deficit patterns that can be used to discriminate neurodegenerative diseases have been published [4, 5]. FDG-PET can reveal pathological changes earlier and more reliably than morphological imaging and even neuropsychological testing [3].

Reliable assessment of FDG-PET images requires a high level of expertise and might be

Table 1. Included subjects

Subject	Diagnosis	Gender	Age	1st dose	2nd dose
1	Control	M	70	314	87
2	Control	M	69	280	72
3	Control	M	78	220	56
4	Control	M	84	298	84
5	Control	M	64	261	63
6	Control	F	70	215	58
7	Control	F	68	240	54
8	Control	F	77	190	45
9	Control	F	78	155	35
10	SD	M	82	264	62
11	bvFTD	M	76	244	49
12	bvFTD	M	56	248	66
13	AD	M	69	227	54
14	bvFTD	F	59	161	35
15	LBD	M	74	207	57
16	CBD	F	84	164	49
17	LBD	M	71	239	47
18	bvFTD	M	74	251	63

Diagnosis (SD = semantic dementia, bvFTD = behavioural variant frontotemporal dementia, AD = Alzheimer's disease, LBD = Lewy body dementia, CBD = corticobasal degeneration). 1st dose represents the normal-dose scan, in MBq. 2nd dose represents the low-dose, also in MBq. A large part of this material was also included in a previously published article [11].

challenging in early stages of disease. Since its introduction in the 1990-s [6], the use of automated 3D-stereotactic spatial normalization maps (3D-SSP maps) has been an increasingly important part of clinical evaluation. The 3D-SSP method defines a large number of points on a spatially normalized brain surface and compares the data from the individual with a database of healthy controls. The result can be used to highlight cortical areas with statistically significant deficits. These deficit maps, referred to as Z-score maps due to the underlying statistical calculations (or simply Z-maps), optimizes the pattern recognition in the reader and facilitate differential diagnosis. FDG-based Z-maps have been shown to improve the accuracy of diagnosing neurodegenerative disease [3, 7], especially for novice readers [8].

There are several commercially available software packages with a 3D-SSP function, e.g. HybridViewer BRASS (Hermes medical solutions), CortexID (GE Healthcare), Neurostat 3D-SSP (University of Michigan, USA), and SPM (Wellcome Dept. London, UK). 3D-SSP maps are commonly used clinically, both for FDG and amyloid imaging, and method optimization is ongoing [9].

During FDG-PET, the subject receives a moderate but not insignificant dose of ionizing radiation [10]. For many patients, this is of negligible importance due to old age and concurrent disease, but for younger patients and for the prospect of repeated examinations, a dose reduction with retained image quality is desirable. With lower dose, repeated examinations become a useful tool for monitoring disease progress and potential effects of a disease-modifying intervention. For research purposes, the possibility to perform low-dose FDG-PET examinations opens up for facilitated inclusion of healthy controls, even for repeated examinations. A normal-dose scan with shorter acquisition time is a close but not perfect approximation of a true low-dose scan. An important difference is a higher count rate, and consequently higher random coincidence rates and dead time during a shortened normal-dose scan, which would negatively affect image quality compared to a true low-dose scan.

It was recently shown that true low-dose FDG-PET provided highly similar regional SUV ratios compared to normal-dose without increasing scan time [11]. However, due to the statistical properties of the Z-score method, and limitations such as database size in the software tools, it remains to be shown that reliable Z-maps can be created from low-dose FDG-PET. In clinical routine, maps showing Z-scores filtered at -1.96 standard deviations are commonly used to show areas of significant deficits. This is considered a robust diagnostic method and the current study is based on assessing such maps. The aim of this study was to evaluate the reliability of 3D-SSP Z-score maps created from low-dose FDG-PET of the brain, compared to normal-dose, using quantitative and qualitative methods.

Material and methods

Subjects

Nine patients (mean age 71.7, 78% males) who had received a clinical probability diagnosis of neurodegenerative dementia disease were prospectively enrolled from the memory clinic at

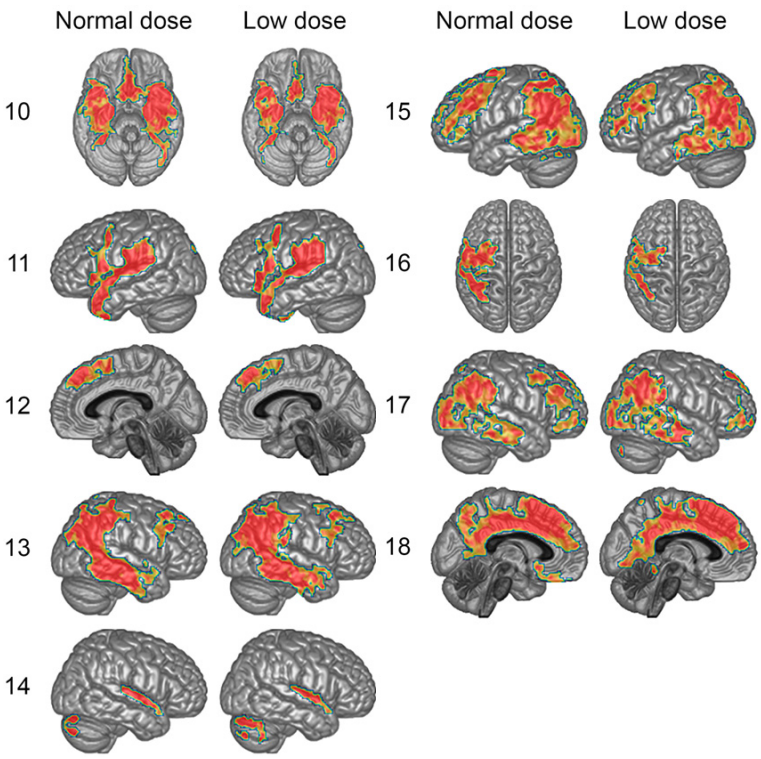

Figure 1. Comparisons of Z-maps from a routine scan with normal dose, and low-dose scan. The figure contains sample images from all included patients in the study (not the controls). The numbers to the left of the images represent the subject numbers. Subject number 16 had the largest difference between normal and low dose according to the assessments described below.

the geriatric department. Five patients had been diagnosed with frontotemporal dementia (behavioural variant (bvFTD), n=4; semantic dementia (SD), n=1, collectively referenced as FTD), two patients with Lewy body dementia (LBD), and one each with Alzheimer's disease (AD) and corticobasal degeneration (CBD), respectively. The diagnoses imply that all patients had neurodegenerative diseases with assumed or previously proven regional hypometabolism, identifiable on FDG-PET.

Nine cognitively normal controls (mean age 71.9, 57% males) were recruited through advertising. A structured interview, physical examination, and neuropsychological tests (MMSE, TMTa, 7-min screening) were performed to exclude pathology and confirm normal cognition. All subjects (n=18) underwent PET/CT twice, as described below. Information about the included subjects is summarized in **Table 1**. The subjects and FDG protocol described here are previously published in greater detail [11, 12].

The study was approved by the Regional Board of Medical Ethics, and all subjects provided written informed consent. For most patients, a present family member co-signed the consent.

FDG-PET protocol

All scans were acquired 35-45 min after tracer injection. A routine scan (normal-dose, ND) with an injected dose of 3 MBq/kg FDG was performed first. The average administered activity was 236 MBq (effective dose from FDG 4.5 mSv). The second scan, performed on a separate occasion, was a low-dose (LD) scan using 25% of the normal dose, i.e. 0.75 MBq/kg (mean administered activity 57 MBq, effective dose from FDG 1.1 mSv.

All subjects were scanned on a Discovery ST (GE Healthcare, Waukesha, WI, USA) PET/CT scanner in 3D acquisition mode. Images were reconstructed using ordered-subsets expectation maximization (OSEM; 2 iterations, 21 subsets) with a 2.14 mm FWHM post-filter, applying appropriate corrections including attenuation correction based on a low-dose CT scan (effective dose 0.07 mSv). Voxel size of the resulting images was 2 × 2 × 3.7. The average time interval between the two examinations was 31 days.

Image post-processing

3D-SSP images showing Z-scores of < -1.96 (henceforth called Z-maps) were created from 10-minute summations (35-45 min) of both ND and LD scans, using batch analysis in HybridViewer BRASS 2.6A.10, Hermes Medical Solutions. Regional Z-scores from 47 regions of interest were generated by the automated atlas function inherent of the BRASS software, using the cerebellum as reference region. The study was comprised of a quantitative part comparing the Z-scores, and a qualitative part where experienced readers visually assessed the 3D-SSP Z-maps, as described below.

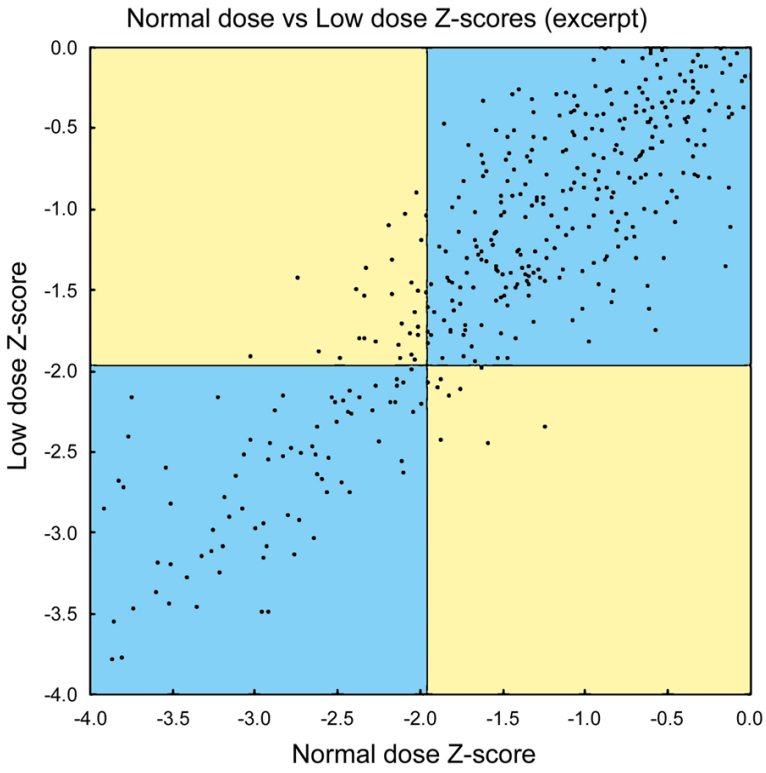

Figure 2. Z-scores from normal-versus low-dose. The plot shows an excerpt of the clinically most relevant Z-score range. Each data point shows one neocortical region in one subject; with the normal-dose derived Z-score on the X-axis and the low-dose derived Z-score on the Y-axis. Using -1.96 as a general cut-off between normal and pathological, the yellow fields contain 41 data points that would be misclassified by low-dose derived Z-scores, compared to normal-dose.

Readers

Two nuclear medicine specialists with vast clinical experience of FDG-PET dementia diagnostics were readers in this study (TD, VI). The images were presented in a random, blinded fashion. The reading was divided into two sessions, performed on separate occasions.

Session one

In the first session of visual assessment, the readers viewed randomized Z-maps from each subject individually, without knowing if the images were a patient or a control, ND or LD. The reader classified each image as normal or pathological, giving a confidence score of 0-5 (discriminative confidence). If the image was classified as pathological, an extra assessment was made regarding the specific diagnosis (AD/FTD/LBD/CBD).

Session two

The second session was carried out on a separate occasion. The readers viewed two sets of Z-maps at a time; the ND and LD images from each subject. In half of the cases (randomly decided), the ND images were above the LD images and vice versa. Selected projections from all included patients are shown in **Figure 1**. The clinical diagnosis of the subject (Control/AD/FTD/LBD/CBD) was given next to the images. For each subject, the readers decided which set of images was most consistent with the clinical diagnosis, together with a subjective quantification of how large the difference between the images were, on a scale from 0-5. Note that a difference of zero, meaning that the images had identical consistency with the diagnosis, was a valid option for the readers as some pairs of images were identical or near identical.

Statistical methods

In the quantitative part of the study, the Z-score differences of all neocortical regions were tested for normal distribution using Shapiro-Wilks test and visualized with normal probability plots for detection of outliers. Single sample t-tests were used to determine if there was constant bias between the normal and low-dose Z-scores, and simple regression was used to quantify proportional bias, expressed as an R^2 and a *p* value.

In the qualitative session one, Cohen's kappa test was used for inter-reader agreement, and to describe the contingency tables, in conjunction to Pearson chi-square. Wilcoxon matched pairs test was used to compare the confidence scores. A proportion test was used for the rate of correct diagnosis. In session two, Wilcoxon matched pairs and mean values with confidence intervals were used to determine if there was any significant preference for ND- or LD-derived images. Graphs and most statisti-

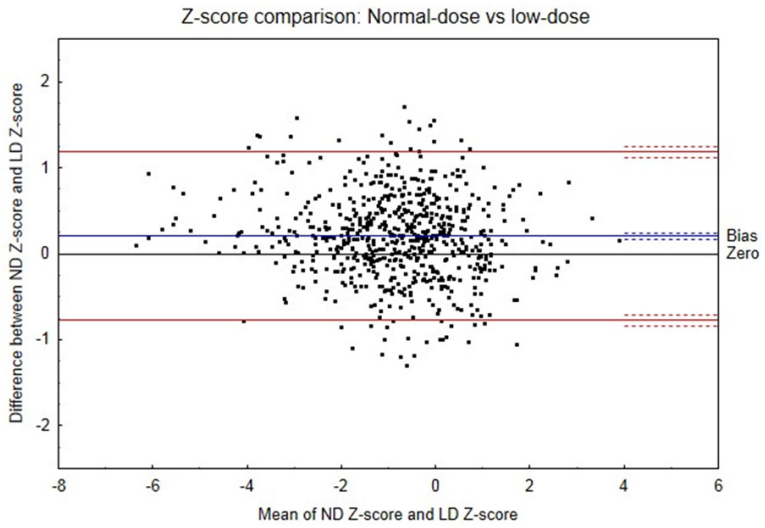

Figure 3. Bland-Altman analysis for neocortical regions, comparing normal- and low-dose. There is a significant constant bias in combination with a slight negative proportional bias, which in combination represents an underestimation of negative Z-scores by the low-dose protocol, compared to normal-dose.

Table 2. Shows the results from the diagnostic assessment in qualitative session 1

	Normal-dose		Low-dose	
Image rated as	Controls	Patients	Controls	Patients
Normal	15	5	15	5
Pathological	3	13	3	13
Of which uncertain	1	1	0	2
Sensitivity	72%		72%	
Specificity	83%		83%	
Negative pred v	75%		75%	
Positive pred v	81%		81%	
Accuracy	78%		78%	
Correct ddx	67%		56%	
	n: 36		n: 36	
	K: 0.556		K: 0.556	
	p: 0.0008		p: 0.0008	

Assessments from both readers are added (n=18 × 2). "Pred v" represents predictive value. The assessments were not identical between normal- and low-dose derived images, but the numbers of assessments in each category were, thus resulting in identical discriminative performance between the methods. "Uncertain" represents assessments with a discriminative confidence score < 3. "Correct ddx" refers to the rate of correct diagnoses among the true positive cases. K represents kappa values for the contingency tables, not to be confused with interreader agreement.

cal calculations were made using Dell Statistica, version 13. Kappa values were calculated with IBM SPSS Statistics version 23.

Results

Selected projections from all included patients are provided in **Figure 1**. The readers were presented with surface projections in all planes but representative images were chosen for this figure.

Quantitative assessment

The difference between the ND- and LD-derived Z-scores were evaluated in several aspects as described in the following sections. Initially, Z-score differences were compared according to the size and signal intensity per voxel in each region of interest, to test for bias from regional size. There was no such significant correlation. A cropped scatter plot of the clinically relevant range of Z-scores is shown in **Figure 2**. Considering -1.96 as a general cut-off limit to separate normal from pathological values, 41 out of 648 (6.33%) data points would be misclassified (false negative, false positive) by the LD protocol compared to ND, as highlighted by the yellow fields in the figure. The differences were normally distributed and no outliers were identified.

Bland-Altman comparisons between the doses showed a significant constant bias of 0.206 ($P < 0.001$), as shown in **Figure 3**. There was also a minimal, but significant, negative proportional bias (R^2= 0.014, P=0.002). The combination of a constant positive bias and a negative proportional bias signifies that negative Z-scores were reported as less negative by the LD protocol compared to ND (e.g. LD showing -2.34 while ND showed -2.54), whereas no such bias was found for positive Z-scores.

This finding was confirmed by dividing the data points to two groups; negative and positive ND-derived Z-scores. The first group had a t-test confirmed significant constant bias, but no bias was found in the second group.

Qualitative assessment

Session one: The results for discrimination between patients and healthy controls are shown in **Table 2**. The sum of assessments in each category was identical, resulting in equal discriminative performance between doses. The LD images had slightly higher mean discriminative confidence score (4.22 versus 4.11 for ND); the ND and LD images were preferred in 11 instances each, thus showing no significant difference. The rate of uncertain cases (discriminative confidence score < 3) was 6% in both dose groups. Kappa values for inter-reader agreement were 0.778 for normal-dose and 0.571 for low-dose. Regarding differential diagnoses, correct diagnoses were given in 67% of normal-dose assessments, and in 56% of low-dose assessments; a non-significant difference at P=0.49.

Session two: When comparing which set of images was most consistent with the clinical diagnosis, the ND images were preferred in 8% of the assessments, the LD images in 13%, and they were deemed equal in 79%. The difference was not significant. The data was also translated to a semi-continuous scale of 11 steps, where "-5" represented strongest difference in favour of LD, "0" represented no difference, and "5" strongest difference in favour of ND. The mean of the assessments on this scale was -0.05, and the 95% confidence interval included zero.

Discussion

The main finding of this study was that visual assessment of 3D-SSP Z-score maps from low-dose FDG-PET provided highly similar diagnostic information, although the quantitative values had some degree of discrepancy compared to normal-dose. The high level of similarity between doses can be of clinical and scientific relevance from a dose-reduction point of view. The findings include a slight understating of negative Z-score values from the LD protocol, and a stronger inter-reader agreement for ND. Regarding differential diagnosis, there was a trend in favour of ND, but no significant difference, and the discrimination between patients and healthy controls were identical between doses. The results from this study in conjunction with the preceding SUV ratio analysis [11] strongly suggest the feasibility of a substantial FDG-dose reduction in clinical routine.

A normal-dose scan with shorter acquisition time approximates a low-dose scan, but there are differences in count rates that adversely affect image quality. Validation using a true low-dose protocol was deemed necessary for clinical implementation, ensuring a valid comparison. In particular, random count rates are proportional to the square of the amount of radioactivity present in the field of view of the scanner, whereas true count rates show a near-linear proportionality at the relatively low activity concentrations encountered during a clinical brain FDG scan. Thus, a scan with a reduced dose will have a higher signal-to-noise ratio than a scan with a reduced acquisition time. As shown in **Figure 1**, the visual similarities between a routine scan and LD were substantial.

In the search for methods to track disease progression rate in clinical trials, many studies have explored the rate of atrophy and the rate of ventricular dilatation, as summarized by Frisoni et al. [13]. Yearly reduction of hippocampal volume has been found to be larger in patients than age-matched controls. Repeated low-dose FDG-PET examinations might provide a similar quantitative "rate-of-progression" measurement for cortical metabolism, which could be useful in identifying pathology at early stages of disease and monitoring disease progression rate [14]. Using 3D-SSP maps to calculate and visualize the rate of progression is likely a robust and appealing approach.

The main limitation of this study is the group sizes. It is theoretically possible that the trend towards superior differential diagnostic capability with ND would have been significant with a larger sample, but visual comparisons of the current material does not support that theory. A second important limitation was the lack of normal-dose reproducibility data based on modern reconstruction algorithms. A separate group of Z-scores using normal-dose FDG-PET/CT twice would provide a reliability measure that might allow stronger interpretations of the

current results. Pairwise comparisons have been shown to be more reliable than subjective grading alone [15], which motivated the second reading session in this study.

It is possible that a scanner with higher sensitivity and higher signal-to-noise ratio would provide slightly different quantitative results, but qualitative assessment would probably remain similar. The data presented in the current work were acquired using a previous generation PET-CT scanner based on BGO detectors and photomultiplier tubes, with considerably inferior image quality than the latest generation of higher-sensitivity, time-of-flight-capable scanners that are equipped with LSO or LYSO detectors and silicon photomultiplier. This, combined with further improvements in image reconstruction algorithms, further strengthens the case for low-dose FDG brain PET.

Conclusion

3D-SSP Z-score maps created from low-dose FDG-PET scans of the brain, using 0.75 MBq/kg, are highly similar to Z-score maps from normal-dose scans.

Acknowledgements

Many thanks to the staff of the radiology department, and the staff of the PET centre, especially Mimmi Lidholm and Lars Lindsjö.

Disclosure of conflict of interest

None.

Address correspondence to: David Fällmar, Department of Surgical Sciences, Radiology, Uppsala University, BFC/Radiology, Uppsala University Hospital, Uppsala 75185, Sweden. Tel: +46730508792; E-mail: david.fallmar@akademiska.se; david.fallmar@gmail.com

References

[1] Prince M, Wimo A, Guerchet M, Ali G, Wu Y, Prina M. World Alzheimer Report 2015. Alzheimer's Disease International (ADI), London.

[2] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ and Scheltens P. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007; 6: 734-746.

[3] Drzezga A. Diagnosis of Alzheimer's disease with [18F]PET in mild and asymptomatic stages. Behav Neurol 2009; 21: 101-115.

[4] Brown RK, Bohnen NI, Wong KK, Minoshima S and Frey KA. Brain PET in suspected dementia: patterns of altered FDG metabolism. Radiographics 2014; 34: 684-701.

[5] Shivamurthy VK, Tahari AK, Marcus C and Subramaniam RM. Brain FDG PET and the diagnosis of dementia. AJR Am J Roentgenol 2015; 204: W76-85.

[6] Minoshima S, Frey KA, Koeppe RA, Foster NL and Kuhl DE. A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET. J Nucl Med 1995; 36: 1238-1248.

[7] Ishii K, Kono AK, Sasaki H, Miyamoto N, Fukuda T, Sakamoto S and Mori E. Fully automatic diagnostic system for early- and late-onset mild Alzheimer's disease using FDG PET and 3D-SSP. Eur J Nucl Med Mol Imaging 2006; 33: 575-583.

[8] Lehman VT, Carter RE, Claassen DO, Murphy RC, Lowe V, Petersen RC and Peller PJ. Visual assessment versus quantitative three-dimensional stereotactic surface projection fluorodeoxyglucose positron emission tomography for detection of mild cognitive impairment and Alzheimer disease. Clin Nucl Med 2012; 37: 721-726.

[9] Lilja J, Thurfjell L and Sorensen J. Visualization and quantification of 3-dimensional stereotactic surface projections for 18F-flutemetamol PET using variable depth. J Nucl Med 2016; 57: 1078-1083.

[10] Huang B, Law MW and Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. Radiology 2009; 251: 166-174.

[11] Fallmar D, Lilja J, Kilander L, Danfors T, Lubberink M, Larsson EM and Sorensen J. Validation of true low-dose (18)F-FDG PET of the brain. Am J Nucl Med Mol Imaging 2016; 6: 269-276.

[12] Fallmar D, Lilja J, Velickaite V, Danfors T, Lubberink M, Ahlgren A, van Osch MJ, Kilander L and Larsson EM. Visual assessment of brain perfusion MRI scans in dementia: a pilot study. J Neuroimaging 2016; 26: 324-330.

[13] Frisoni GB, Fox NC, Jack CR Jr, Scheltens P and Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 2010; 6: 67-77.

[14] Shokouhi S, Claassen D, Kang H, Ding Z, Rogers B, Mishra A, Riddle WR; Alzheimer's Disease Neuroimaging Initiative. Longitudinal progression of cognitive decline correlates with changes in the spatial pattern of brain 18F-FDG PET. J Nucl Med 2013; 54: 1564-1569.

[15] Phelps AS, Naeger DM, Courtier JL, Lambert JW, Marcovici PA, Villanueva-Meyer JE and MacKenzie JD. Pairwise comparison versus Likert scale for biomedical image assessment. AJR Am J Roentgenol 2015; 204: 8-14.