

FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease

Norman L. Foster,¹ Judith L. Heidebrink,^{2,4} Christopher M. Clark,⁵ William J. Jagust,⁷ Steven E. Arnold,^{5,6} Nancy R. Barbas,^{2,4} Charles S. DeCarli,⁸ R. Scott Turner,^{2,4} Robert A. Koeppe,³ Roger Higdon⁹ and Satoshi Minoshima¹⁰

¹Center for Alzheimer's Care, Imaging and Research and Department of Neurology, University of Utah, ²Department of Neurology, University of Michigan, ³Division of Nuclear Medicine, University of Michigan, ⁴Ann Arbor Veterans Administration Hospital, ⁵Alzheimer's Disease Center, Institute on Aging and Department of Neurology, University of Pennsylvania, ⁶Department of Psychiatry, University of Pennsylvania, ⁷Department of Neuroscience, University of California, ⁸Department of Neurology, University of California at Davis, ⁹BIATECH Institute and ¹⁰Department of Radiology, University of Washington, USA

Correspondence to: Norman L. Foster, MD, 650 Komas Drive, #106-A, Salt Lake City, UT 84108-1225, USA
E-mail: norman.foster@hsc.utah.edu

Distinguishing Alzheimer's disease (AD) and frontotemporal dementia (FTD) currently relies on a clinical history and examination, but positron emission tomography with [¹⁸F] fluorodeoxyglucose (FDG-PET) shows different patterns of hypometabolism in these disorders that might aid differential diagnosis. Six dementia experts with variable FDG-PET experience made independent, forced choice, diagnostic decisions in 45 patients with pathologically confirmed AD (n = 31) or FTD (n = 14) using five separate methods: (1) review of clinical summaries, (2) a diagnostic checklist alone, (3) summary and checklist, (4) transaxial FDG-PET scans and (5) FDG-PET stereotactic surface projection (SSP) metabolic and statistical maps. In addition, we evaluated the effect of the sequential review of a clinical summary followed by SSP. Visual interpretation of SSP images was superior to clinical assessment and had the best inter-rater reliability (mean kappa = 0.78) and diagnostic accuracy (89.6%). It also had the highest specificity (97.6%) and sensitivity (86%), and positive likelihood ratio for FTD (36.5). The addition of FDG-PET to clinical summaries increased diagnostic accuracy and confidence for both AD and FTD. It was particularly helpful when raters were uncertain in their clinical diagnosis. Visual interpretation of FDG-PET after brief training is more reliable and accurate in distinguishing FTD from AD than clinical methods alone. FDG-PET adds important information that appropriately increases diagnostic confidence, even among experienced dementia specialists.

Keywords: Alzheimer's disease; PET; FDG; frontotemporal dementia

Abbreviations: AD = Alzheimer's disease; FDG = fluorodeoxyglucose; FTD = frontotemporal dementia

Received February 21, 2007. Revised July 6, 2007. Accepted July 11, 2007. Advance Access publication August 18, 2007

Identifying the specific cause of dementia is challenging and increasingly important as effective, disease-specific treatments have become available. Clinicians particularly need a practical method to accurately differentiate frontotemporal dementia (FTD) from Alzheimer's disease (AD). The first symptom of AD is typically memory loss, while the hallmarks of FTD are behaviour and language disturbance (Kertesz and Munoz, 1998). However, both disorders cause an insidious, gradually progressive dementia that lacks distinctive physical signs, and patients with FTD frequently meet diagnostic criteria for AD (Varma *et al.*, 1999). Thus, it is not surprising that FTD is frequently misdiagnosed, even in specialty clinics (Mendez *et al.*, 1993).

It is important for physicians to determine whether AD or FTD is the cause of dementia. FTD is a common cause of early-onset dementia and in the 45 to 64-year age group, AD and FTD have an equal prevalence of 15 per 100 000 (Ratnavalli *et al.*, 2002). A diagnosis of FTD can have significant implications for family members. Approximately one-third of patients with FTD have a family history of a similar disorder, and relatives are at increased risk of dementia at an earlier age than the general population (Rosso, 2003). FTD and AD have different pathology and prognosis. In FTD, causative genetic mutations and histology indicate a disturbance in the microtubule protein tau or the trophic factor progranulin, instead of in the

beta-amyloid, as is characteristic of AD (Hardy and Selkoe, 2002; Trojanowski and Mattson, 2003). FTD lacks the cholinergic deficiency of AD and its distinctive symptoms and clinical course often present different management challenges (Procter *et al.*, 1999; Foster, 2003a). These discrepancies explain why the appropriate treatment of AD and FTD differ and why these disorders need to be distinguished clinically.

Positron emission tomography with [¹⁸F] fluorodeoxyglucose (FDG-PET) highlights the different distribution of pathology in dementing disorders and might aid diagnosis. Recognizing AD and FTD is a particularly promising application for FDG-PET because of the sharp contrast in their pattern of glucose hypometabolism. AD causes hypometabolism predominantly in posterior regions: the posterior temporoparietal association cortex and posterior cingulate cortex (Minoshima *et al.*, 1997). FTD causes hypometabolism predominantly in anterior regions: the frontal lobes, anterior temporal cortex and anterior cingulate cortex (Ishii *et al.*, 1998). Although FDG-PET has been used to study neurodegenerative disease for over two decades, its diagnostic potential has not been fully exploited. Most studies have been designed to understand the biology of dementia and are inadequate to assess clinical utility (Gill *et al.*, 2003). Evaluation of a diagnostic test relies upon individual, rather than group differences from a reference population and is assessed with statistical measures such as sensitivity, specificity, predictive value and likelihood ratio. These measures apply to a single diagnostic comparison. A history suggesting cognitive impairment confirmed on clinical examination is the best and least costly way to distinguish between normal and demented patients, but greater clinical experience and judgement are needed to distinguish different kinds of dementing diseases. Now that FDG-PET is becoming widely available, clinical trials to evaluate whether FDG-PET has an important role in the evaluation of dementia are timely and necessary.

Methods

Overview

We evaluated the utility of FDG-PET to distinguish AD and FTD in individual patients whose diagnoses were known from histopathological examinations using methods easily incorporated in clinical practice. Initially we compared five separate diagnostic procedures: three entailing review of only clinical information and two involving review of only FDG-PET imaging (Table 1). We then examined the sequential use of the most accurate clinical method and most accurate imaging method. This permitted us to determine whether FDG-PET provided any added benefit in a dementia evaluation. Six raters independently used each diagnostic approach to assign a diagnosis of AD or FTD. For each diagnostic approach we determined its reliability, characteristics as a diagnostic test and effect on diagnostic confidence.

Subjects

We identified all patients with dementia who had an FDG-PET scan at the University of Michigan between December 1984 and

Table 1 Diagnostic methods evaluated in this study

Independent diagnostic methods	
Clinical data:	<ol style="list-style-type: none"> 1. Interpretation of clinical scenario 2. Symptom checklist score 3. Interpretation of clinical scenario with symptom checklist score
FDG-PET imaging data:	<ol style="list-style-type: none"> 1. Transaxial images of glucose metabolism relative to pons 2. Stereotactic surface projection (SSP) metabolic and statistical maps
Sequential diagnosis	
<ol style="list-style-type: none"> 1. Interpretation of clinical scenario alone 2. Interpretation of clinical scenario also considering SSP metabolic and statistical maps 	

July 1998 and subsequently received a post-mortem examination documenting a histopathological diagnosis of AD or FTD, uncomplicated by other pathology such as stroke or significant numbers of cortical Lewy bodies. Only individuals with retrievable parametric PET images that included most of the brain in the field of view were considered. Of the total 48 individuals found in our record review, three were excluded because their medical records were not retrievable. Of the remaining 45 patients, 31 had definite AD and 14 had FTD. AD patients met NIA–Reagan neuropathological criteria for either high (28 cases) or intermediate (3 cases) likelihood of AD (NIA and Reagan Institute Working Group, 1997). We identified minor additional pathological abnormalities in eight of these subjects: three with cortical Lewy bodies insufficient to meet neuropathological criteria for dementia with Lewy bodies (McKeith *et al.*, 1996), and five with cortical arteriolosclerosis, including two with subcortical lacunar infarctions of indeterminate age. Two AD subjects, ages 34 and 35 years at the time of their scans, had early-onset familial AD. Excluding these two individuals, the mean age of AD subjects was 67.8 ± 7.6 (age range 51–79 years).

Patients with FTD had several specific neuropathological diagnoses generally recognized as causing the clinical syndrome of frontotemporal dementia, including frontotemporal degeneration without distinctive histopathology (five cases), Pick's disease (four cases), corticobasal degeneration (two cases), progressive subcortical gliosis (one case), mesocorticolimbic degeneration (one case) and frontotemporal dementia with parkinsonism linked to chromosome 17 and a mutation in the TAU gene (FTDP-17T) (one case). The presence or absence of progranulin mutations was not assessed and we did not apply recently developed ultra-sensitive ubiquitin antibodies in these cases.

Subjects were identified from autopsy results rather than clinical diagnoses. Several patients with AD pathology had an atypical presentation with prominent language or visual symptoms (Table 2). One patient with AD had a particularly rapid course and was clinically thought to have Creutzfeldt–Jakob disease. Another was thought to have Parkinson's disease with dementia. Individuals with FTD pathology were not prospectively classified into clinical subtypes because their initial evaluations occurred between 1985 and 1998, and only two occurred after the first clinical FTD criteria were published in 1994. As a result, except for

Table 2 Characteristics of individual study subjects

Case number	Pathologic diagnosis	Age at Sx onset	Duration of Sx at first visit (yrs)	First visit to PET (yrs)	PET to death (yrs)	Clinical diagnosis and presentation	Retrospective consensus diagnosis
1	AD	67	5	0.0	4	AD	AD
2	AD	59	3	0.3	3	Suspected Creutzfeldt-Jakob disease	FTD, CBD
3	AD	76	2	0.8	4	AD, prominent aphasia	AD
4	AD	72	3	0.1	3	AD	AD
5	AD	65	6	0.1	4	AD	AD
6	AD	64	6	0.0	5	AD	AD
7	AD	57	2	1.6	9	AD, prominent aphasia	FTD, mixed semantic/PNFA
8	AD	32	1	2.3	4	AD	AD
9	AD	74	3	0.0	3	AD	AD
10	AD	47	5	2.0	1	AD	AD
11	AD	72	5	1.8	8	AD, atypical slow course	AD
12	AD	68	3	0.4	6	AD	AD
13	AD	61	3	3.8	3	AD, prominent aphasia	AD
14	AD	55	1	1.0	7	Parkinson's disease with dementia	AD
15	AD	46	5	0.1	5	AD	AD
16	AD	74	2	0.4	3	AD	AD
17	AD	33	1	0.0	2	AD	AD
18	AD	49	15	0.2	7	AD	AD
19	AD	54	3	0.7	3	AD, atypical, possible CBD	AD
20	AD	65	4	1.6	3	AD, prominent visual	AD
21	AD	64	2	0.5	5	AD	AD
22	AD	57	4	0.1	7	AD	AD
23	AD	62	7	2.3	4	AD	AD
24	AD	58	7	2.6	5	AD, prominent aphasia	AD
25	AD	64	4	0.3	6	AD, prominent aphasia	FTD, CBD
26	AD	69	4	4.1	4	AD, atypical slow course	AD
27	AD	65	3	0.6	5	AD	FTD, mixed behavioural/PNFA
28	AD	58	3	0.1	6	AD	AD
29	AD	58	3	0.2	5	AD, prominent visual	AD
30	AD	69	6	0.1	5	AD	AD
31	AD	68	4	1.0	6	AD	AD
32	FTD	74	6	1.2	4	AD	AD
33	FTD	58	2	1.5	6	AD, prominent aphasia	AD
34	FTD	60	1	2.3	9	AD, prominent aphasia	FTD, semantic
35	FTD	58	1	0.0	2	PSP	FTD, PSP
36	FTD	60	2	0.4	11	AD, atypical frontal	FTD, behavioural
37	FTD	59	2	2.7	6	AD, atypical frontal	FTD, semantic
38	FTD	58	4	2.5	3	AD	FTD, behavioural
39	FTD	57	7	0.5	5	AD	AD
40	FTD	66	1	2.2	3	AD, atypical frontal	FTD, PSP
41	FTD	60	1	0.0	8	AD, atypical frontal	FTD, behavioural
42	FTD	65	2	0.1	5	AD, atypical frontal	FTD, mixed behavioural/PNFA
43	FTD	54	10	0.3	0	AD, atypical frontal	FTD, behavioural
44	FTD	63	6	0.1	7	AD, prominent aphasia	FTD, behavioural
45	FTD	59	10	0.1	0	AD, atypical frontal	FTD, mixed behavioural/PNFA

Note: AD: Alzheimer's disease, FTD: frontotemporal dementia, PNFA: progressive non-fluent aphasia, PSP: progressive supranuclear palsy, CBD: corticobasal degeneration.

one individual diagnosed with progressive supranuclear palsy, all subjects with FTD histopathology received an initial clinical diagnosis of AD. Nevertheless, medical records indicate seven presented primarily with frontal symptoms of personality change and behaviour disturbance and three presented with predominant aphasia. A separate panel of six dementia specialists also provided

a consensus diagnosis based upon their retrospective review of the clinical scenarios, knowing that pathology showed either AD or FTD (Table 2). For cases diagnosed as FTD, they also provided a subtype classification based upon published guidelines (Neary *et al.*, 1998). Several were difficult to classify into a single category and had features of more than one subtype.

Table 3 Summary characteristics of study subjects

Diagnosis	AD (n = 31)	FTD (n = 14)	All patients (n = 45)	Controls (n = 33)
Prevalence in this study	69%	31%	100%	–
Gender	20 men 11 women	7 men 7 women	27 men 18 women	19 men 14 women
Age at scan	65.6 ± 11.1 (range 34–79)	65.6 ± 5.5 (range 59–81)	65.6 ± 9.6 (range 34–81)	68.5 ± 8.2 (range 58–91)
MMSE at scan	14.0 ± 8.7 (n = 25) (range 0–27)	15.5 ± 9.5 (n = 10) (range 0–24)	14.4 ± 8.8 (n = 35) (range 0–27)	N/A
Time from symptom onset to first clinic visit (years)	4.0 ± 2.6 (range 1–15)	3.9 ± 3.3 (range 1–10)	4.0 ± 2.8 (range 1–15)	–
Time from first clinic visit to PET scan (years)	0.9 ± 1.1 (range 0–4.1)	1.0 ± 1.0 (range 0–2.7)	1.0 ± 1.1 (range 0–4.1)	–
Time from PET scan to death	4.7 ± 1.8 (range 1–9)	4.9 ± 3.2 (range 0–11)	4.7 ± 2.3 (range 0–11)	–
Scan dates	12/5/84–1/11/95	8/30/85–7/2/98	12/5/84–7/2/98	7/15/93–8/10/99
Scanner type	3 EXACT 9 TCC 19 ECAT	1 EXACT 4 TCC 9 ECAT	4 EXACT 13 TCC 28 ECAT	32 EXACT 0 TCC 1 ECAT

Note: Values are mean ± SD; N/A = not available; EXACT = Siemens/CTI Exact 47 scanner (xy pixel dimension 1.91 mm, in-plane × axial resolution 8.0 × 5.0 mm; TCC = The Cyclotron Corporation PCT 4600a scanner (xy pixel dimension 3.75 mm, in-plane × axial resolution 12.0 × 9.5 mm; ECAT = Siemens/CTI ECAT 931 scanner (xy pixel dimension 1.89 mm, in-plane × axial resolution 8.0 × 7.5 mm).

AD and FTD subjects had similar demographic characteristics (Table 3). Initial evaluations in our clinic occurred on average 4 years after symptom onset, although sometimes symptoms reportedly had been present for a decade or more and dementia was already severe. Two-thirds of subjects had their FDG-PET scan within 1 year of their first visit.

We also identified 33 cognitively normal elderly individuals of similar age to our study subjects who had received FDG-PET scans as control subjects for previous research studies (Table 3). We constructed a database of scans from these control subjects for statistical comparisons with patient scans.

Raters

Six neurologists with 10 to 25 years of experience in dementia care at three NIA-funded Alzheimer's Disease Centers served as raters (SA, NB, CC, CD, WJ and RST). FDG-PET research studies had been conducted at all three Centers, but the raters themselves had variable imaging experience; some were recognized experts in FDG-PET imaging, others were novices. We selected two raters from each Center so regional and institutional differences could be examined. All raters were informed that study subjects had an autopsy-confirmed diagnosis of either FTD or AD, but they did not know the proportion of subjects with each diagnosis. Institutional Review Boards at the University of Michigan and at each of the investigator's institutions approved this study.

Clinical scenarios

We developed summaries extracted from all available medical records of the clinical course of each patient. Often, serial dementia clinic assessments performed over many years were available, and many patients were followed until shortly before death. A research assistant redacted all personal identifiers, clinical diagnoses, the results of imaging studies, genetic analyses and autopsy reports. All subjects received structural imaging studies,

either CT or MRI, as part of their clinical assessment. None showed focal lesions. Structural imaging studies and detailed neuropsychological data were redacted from the materials used to generate the case scenarios to reduce bias. Many scans were not retrievable for direct review and their methods and the quality of the reports varied considerably. Likewise, neuropsychological testing was inconsistent and therefore we only provided summary scores. A single neurologist (JH), experienced in dementia assessments and unaware of subject identity, diagnosis and pathologic findings, reviewed the redacted medical records to develop a chronological summary of the patient's entire illness, averaging 650 words in both AD and FTD subjects. These clinical scenarios focused on the patient's initial and most prominent symptoms, and the results of mental status and neurological examinations. They included illustrative examples of symptoms and the results of neuropsychological testing when available (see example, Appendix 1). They were similar in length and content to summaries used in another study of diagnostic reliability and validity in dementia, except they did not include diagnostic imaging results (Blacker *et al.*, 1994). Case scenarios were assigned random numbers and sent to the raters for review. Based solely upon the scenarios, raters were asked to make a diagnosis of AD or FTD and indicate their degree of diagnostic confidence—very confident, somewhat confident or uncertain.

Diagnostic checklist

Symptom checklists have been advocated as an aid in diagnostic decision-making. After the raters used the clinical scenario to reach a diagnosis, they were asked to use the clinical scenario to complete and score a 26-item questionnaire developed by Barber *et al.* (1995). This checklist identifies symptoms felt to be characteristic of either AD or FTD based upon the timing of their appearance in the course of disease. We followed the rules outlined by Barber *et al.* to complete the questionnaire, but made

minor formatting and grammar modifications to accommodate our use of the questionnaire in a record review, rather than in its original design as an informant interview. The rater divided the patient's clinical course into thirds and then determined whether specific symptoms identified in the checklist were present or absent. For example, in the first third of the illness a change in personality increases the score to favour FTD, while geographic disorientation and learning problems decreases the score to favour AD. When raters thought information in the scenarios was insufficient to assess a specific stage of the illness, the section was omitted. However, even in those circumstances, the checklist still provided a score that could be translated into a diagnosis using the scoring rules (positive scores and zero indicated FTD, negative scores indicated AD). The results of the checklist were used in two ways. First, we simply recorded the diagnosis indicated by the checklist score. Second, raters were asked to make a diagnosis of either AD or FTD and indicate their degree of confidence after completing the checklist and considering the computed score along with the clinical scenario.

Image processing

FDG-PET data were obtained from archived files. PET instrumentation and methods for reconstructing parametric images of glucose metabolism have evolved rapidly over the years. Thus, procedures for attenuation correction, scatter correction and filtering were different depending on when the scan was performed. We obtained FDG-PET scans from three PET instruments with bismuth germanium oxide detectors. These scanners have different technical specifications and resolutions and use different data file formats. Fortunately, our archiving system was able to retrieve and manipulate scan files from all these instruments, despite the evolution in imaging acquisition that occurred over the years. All scans in this study included the entire brain including the brainstem, a requirement of our image analysis software. Although it has a small axial field of view, this was achieved in the TCC scanner using a series of 2–4 contiguous and interleaved scans, each consisting of five transaxial images. The use of multiple PET instruments could be a major concern in research studies based upon quantitative image analysis. However, we assumed that metabolic changes due to disease were likely greater than those due to variations in FDG-PET data acquisition.

We provided raters with two different colour displays of FDG-PET scan data—transaxial and stereotactic surface projection (SSP) images. Because variations in FDG-PET data acquisition may affect topographic patterns of glucose hypometabolism, we wanted to investigate whether a voxel-wise method like SSP would aid visual inspection of data from PET instruments with varying physical specifications. Images had no personal identifiers or dates and were labelled with randomly generated numbers. Transaxial and SSP images and clinical scenarios were sent to raters on separate dates and had different random number labels. Thus, raters could not compare transaxial and SSP images from the same subject or compare PET images to clinical scenarios.

Raters received all relevant transaxial images available for each subject (15–47 per subject, depending upon the scanner) in a standard format and orientation. Images were shown as relative metabolic rates with the highest pixel value in the scan placed at the highest value on the colour scale (Fig. 1A).

Traditionally, physicians have viewed FDG-PET scans as a series of transaxial images. The first PET scanners produced only a single

'slice' image. Multi-slice PET instruments have higher spatial resolution and provide more detailed information. The large number of transaxial images these instruments generate (128 or more in some current models) presents a challenge to the clinician. The images must be mentally manipulated into a 3D space to recognize, describe and interpret a metabolic pattern. SSP is an automated analysis method that warps images into a uniform stereotactic space and also permits statistical analysis of individual scans. Each set of brain images is first oriented along a line passing through the anterior and posterior commissures. Then, through a series of automated steps, imaging data are interpolated to establish a uniform image matrix and voxel size. Next, linear scaling is used to correct for individual brain size and regional anatomic differences with the Talairach atlas brain are minimized with non-linear warping. This enables reliable pixel-by-pixel comparisons of these anatomically standardized brain images.

SSP is designed to select data relevant for the interpretation of scans in diseases primarily affecting the cerebral cortex and to summarize this information in a series of 6 easily interpreted surface projection maps (Minoshima *et al.*, 1995b). To determine projection map values, it uses a predetermined vector that is 6 pixels (13.5 mm) long and oriented perpendicular to the outer and medial surfaces of the right and left-brain hemispheres for each surface pixel. The surface pixel is assigned the highest pixel value found along this vector. Because SSP selects peak rather than average for analysis it is relatively resistant to affects of atrophy (Ishii *et al.*, 2001). Previous studies suggest it may have higher reliability for diagnostic decision-making than transaxial images (Burdette *et al.*, 1996). We normalized surface pixel values to the pons, which is relatively preserved in AD (Minoshima *et al.*, 1995a). We determined pons activity by averaging the highest 300 pixels within the pontine region encompassed by the ventrodorsal levels between -16 and -36 mm below the anterior–posterior commissural line.

SSP results are displayed as true surface maps rather than a transparent view of the brain surface often used in other techniques. Raters received two complementary sets of SSP images. The first, a metabolic map, shows values of glucose metabolism relative to pons using the same colour scale as the transaxial images. The second, a statistical map, shows surface pixel-by-pixel z-scores derived from comparing an individual's scan with results in normal controls. The statistical map shows only pixels with significant glucose hypometabolism compared to the control population using a colour scale reflecting the degree of significance. Figure 1B shows an example of both SSP image displays for the same subject shown in Fig. 1A.

Rater training for FDG-PET interpretation

FDG-PET images in this study were evaluated only after raters completed a 2-h training session designed to reduce the impact of their varied imaging expertise and to establish a uniform approach to interpretation. For this training, we developed a set of FDG-PET images not otherwise used in our study from 10 subjects with clinically diagnosed AD, 10 with clinically diagnosed FTD and 5 normal elderly controls. We selected these images to illustrate the full range of findings the raters might expect to encounter in each diagnostic group. The images were labelled only with a consecutive number and diagnosis, except for normal subjects, where age also was provided. Raters were trained via telephone while viewing images on their personal computers.

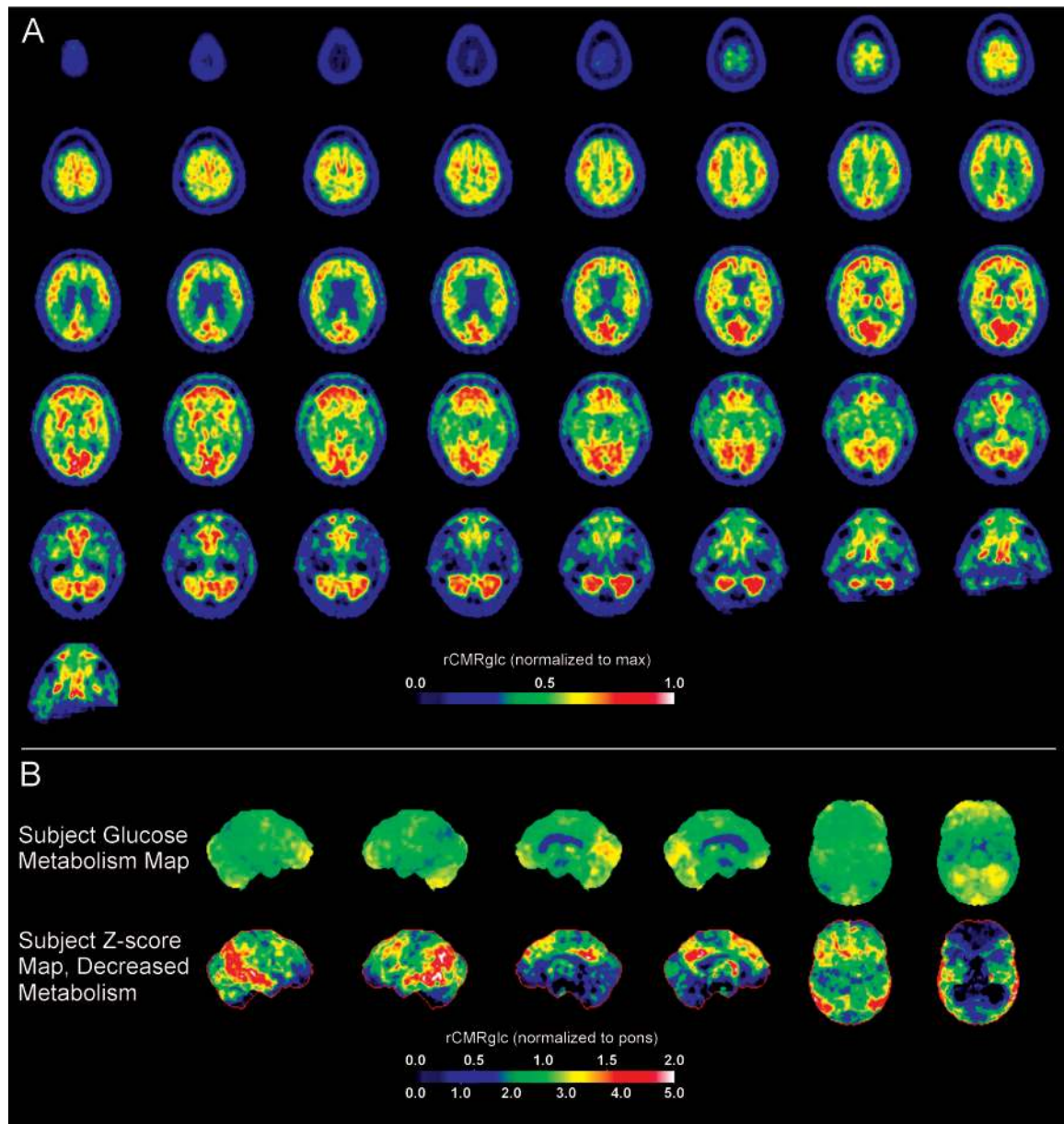


Fig. 1 Example of transaxial and SSP images from a patient with AD separately and independently provided to raters. **(A)** The top six rows of images are a set of 41 transaxial images extending from the top (top left) to the bottom of the brain. The posterior part of the brain in the last images is outside the scanner field of view. **(B)** The lower two rows show stereotactic surface projection (SSP) maps of glucose metabolism relative to pons and pixels with significant z-scores compared to 33 elderly normal control subjects (last row). SSP maps provide six views of the brain—right and left lateral, right and left medial, superior and inferior (in order from left to right of the figure). Values in all images are shown in a colour scale with values red > yellow > green > blue as indicated on the colour bar. Colour bars and labels are provided for reference, but were not included in the images sent to the raters.

The training session began with a review of PET methodology, including technical issues that determine image quality, methods of image display and factors that affect image appearance. Next, imaging anatomy was reviewed with particular attention to regions affected in FTD and AD. Most of the session was spent reviewing and discussing images in the training set. Raters were asked to evaluate training images using the same procedures they would use later with study images. First, raters had to grade the degree of overall scan abnormality as normal, uncertain abnormal, somewhat abnormal or very abnormal. The rating of uncertain

abnormal encouraged raters to identify even the mildest abnormalities in the scan. Second, by focusing attention on the areas most critical to the diagnosis of AD and FTD, they had to decide whether metabolism was normal or abnormal. Five specific brain regions were rated in each hemisphere so a total of 10 regions were assessed. These areas were the posterior temporoparietal association cortex, posterior cingulate gyrus, frontal association cortex, anterior temporal cortex and anterior cingulate gyrus. Since rating the entire scan considered all regions, and not just those rated individually, it was possible for the entire

scan to be rated abnormal, even if all of these 10 regions were considered normal. Third, raters had to decide whether there was significant asymmetry in the degree of hypometabolism in left and right cerebral hemispheres. Finally, raters had to make a diagnosis of either FTD or AD and indicate their degree of diagnostic confidence. Raters were instructed to use simple rules to assign a diagnosis from the images. They were asked to interpret a scan as AD when the degree of hypometabolism appeared greater in the posterior association cortex and posterior cingulate gyrus than in the anterior regions, and as FTD when hypometabolism appeared greater in the frontal association cortex, anterior temporal cortex and anterior cingulate gyrus than in the posterior regions. Raters were forced to make a diagnosis of FTD or AD in each case, even if they rated the scan as normal.

Sequential assessment of clinical scenarios and FDG-PET

Analysis of results of the three clinical and two FDG-PET rating methods found that the diagnostic checklist had no appreciable effect on diagnostic accuracy and that SSP was superior to transaxial image display for FDG-PET. Consequently, after completing these initial ratings, we used only clinical scenarios and SSP FDG-PET images in a second round of ratings to evaluate the potential value of adding FDG-PET to clinical evaluations. For this part of the study, rating of the scenarios and SSP images was accomplished using a web-based system that assured incremental data presentation. No personal identifiers were included and we used a random code number different from those used in previous ratings. First, raters reviewed the case scenario and entered their diagnosis and degree of diagnostic certainty. Only after this response was confirmed and locked were raters allowed to view the subject's SSP images. To ensure that raters were appropriately attentive to the scan, we again asked them for assessments of the scan as a whole, the critical five brain regions in each hemisphere relevant to our diagnostic rules, and the presence of hemispheric asymmetry. After completing these assessments, raters again were asked to assign a diagnosis of FTD or AD and indicate their degree of confidence, this time considering both the FDG-PET scan and scenario, which was still available for review on their computer screen.

Statistical analysis

We compared the diagnostic judgements of raters to neuropathological diagnosis, which served as our reference standard. Since six expert clinicians performed independent assessments on 45 subjects, there were 270 observations for each measure in the study. Inter-rater reliability was assessed using rater agreement and kappa statistics calculated for all possible rater pairs. Rater consensus was evaluated using both unanimity and supermajority (agreement of 5/6 raters) rules. The degree of agreement based upon kappa statistics was rated as fair (kappa values 0.2–0.39), moderate (kappa 0.4–0.59), substantial (0.6–0.79) or almost perfect (0.8–1.0), according to convention (Landis and Koch, 1977). Diagnostic accuracy was assessed by computing rater consensus, the proportion of ratings with correct diagnoses, and standard methods for assessing a diagnostic test—sensitivity, specificity, predictive values and likelihood ratios (Qizilbash, 2002). Because raters had only two diagnostic options, sensitivity and specificity for FTD was equivalent to that for AD and positive and negative predictive values were complementary. We analysed

rater performance by comparing diagnostic accuracy and confidence and their change with addition of PET using graphic displays and statistical tests for binary data. Logistic regression models were fit to binary variables representing whether the diagnosis was correct (accuracy, sensitivity, specificity and predictive value), whether the rater was 'very confident' in their diagnosis, or whether a change in confidence was appropriate for a given diagnosis. A different logistical model accounting for ratios was applied to determine whether there were significant differences in likelihood ratios. Some tests were conditional on the diagnosis (FTD or AD) or whether a change in diagnosis or confidence occurred. Since ratings of the same case by different raters or the same rater using a different method and ratings of different cases by the same rater are potentially correlated, standard independence assumptions do not hold. Statistical tests comparing different diagnostic methods adjusted standard errors and hypothesis tests to account for correlations between cases and raters. This adjustment used a robust variance estimate that incorporates estimates of correlation from these two sources (Andrews, 1991). We then used this adjusted variance estimate to generate *P*-values based on Wald tests.

Results

FDG-PET findings

Raters found almost all scans abnormal (256/270, 95% for transaxial; 267/270, 99% for SSP). In a small portion, the degree of abnormality was mild and considered uncertain. Most scans were rated as either somewhat or very abnormal (84% of transaxial and 95% of SSP scans). Scans from AD and FTD subjects had similar degrees of abnormality. Hypometabolism was more frequent in frontal, anterior cingulate and anterior temporal regions in FTD, and in temporoparietal and posterior cingulate regions in AD, consistent with the criteria used in this study to interpret FDG-PET scans. However, each of these regions was sometimes rated as hypometabolic in both AD and FTD subjects. Likewise, none of the regions was rated as abnormal in all patients with AD or in all patients with FTD. Raters found posterior cingulate hypometabolism in a much higher proportion of subjects with SSP than transaxial images (71 versus 32% in the AD subjects). Otherwise, the proportion of subjects with hypometabolism in a particular region was similar with the two methods.

Significant hemispheric metabolic asymmetry was present in approximately half of both AD and FTD cases, with very similar results using either transaxial or SSP images. A similar proportion of FTD cases had left or right-hemispheric asymmetry, while the right hemisphere was more often hypometabolic in the AD subjects with significant asymmetry (34% rated predominantly right hemisphere hypometabolism and 17% rated predominantly left hemisphere hypometabolism).

Inter-rater reliability

The diagnostic agreement between raters was higher for both FDG-PET methods than for any of the purely clinical

Table 4 Diagnostic inter-rater reliability of five independent methods to distinguish AD and FTD in 45 autopsy-confirmed subjects

	Clinical scenario	Symptom checklist	Scenario + checklist	Transaxial FDG-PET	SSP FDG-PET
All raters agree on diagnosis	23/45	20/45	21/45	36/45	37/45
5/6 raters agree on diagnosis	32/45	37/45	33/45	40/45	42/45
Mean inter-rater kappa	0.42	0.31	0.42	0.73	0.78
(range of individual raters)	(0.25–0.54)	(0.06–0.66)	(0.25–0.57)	(0.53–0.94)	(0.65–0.94)

Table 5 Diagnostic accuracy, sensitivity and specificity of five independent diagnostic methods

	Clinical scenario	Symptom checklist	Scenario + checklist	Transaxial FDG-PET	SSP FDG-PET
All raters correct diagnosis	22/45	20/45	20/45	32/45	37/45
5/6 raters correct diagnosis	29/45	34/45	30/45	36/45	38/45
Mean diagnostic accuracy	78.8%	76.3%	79.6%	84.8%	89.2%
(range of individual raters)	(73–87)	(71–89)	(73–89)	(82–87)	(87–91)
Mean FTD sensitivity/AD specificity	63%	37%	62%	59%	73.2%
(range of individual raters)	(36–79)	(7–79)	(36–79)	(43–71)	(57–82)
Mean FTD specificity/AD sensitivity	86%	94%	88%	96%	97.6%
	(74–100)	(74–100)	(74–100)	(92–100)	(94–100)

methods (Table 4). Transaxial and SSP FDG-PET images showed substantial inter-rater diagnostic agreement based on mean kappa values, and agreement was slightly but not significantly higher for SSP than for transaxial scans. The review of clinical scenarios had moderate inter-rater diagnostic agreement, and there was only fair inter-rater agreement on the diagnosis based upon the symptom checklist score. Although raters were asked to assign a diagnosis again after completing the checklist, their responses were rarely changed from those after the scenario alone and inter-rater agreement therefore was similar.

Although diagnostic judgements in this study were made on the basis of a global assessment of the pattern of hypometabolism, an algorithmic approach using independent judgements about specific brain region abnormalities is a feasible alternative approach. Therefore, we examined the inter-rater reliability and frequency of reported abnormality of individual regions. There was less inter-rater agreement about whether a specific brain region was hypometabolic than about diagnosis, although it was consistently greater in every region with SSP than with transaxial images. Mean kappa values for individual regions ranged from 0.14 to 0.51 using transaxial images and from 0.36 to 0.74 using SSP images. Reliability was not dependent upon the degree of reported hypometabolism. SSP improved inter-rater reliability most in the posterior cingulate cortex while increasing the proportion of AD subjects with hypometabolism recognized in this region (mean kappa 0.20 for transaxial images and 0.57 for SSP). SSP also improved mean kappa scores for the anterior temporal cortex as the proportion with hypometabolism in this region declined. Raters had more agreement on the significance and direction of hemispheric metabolic

asymmetry with transaxial images (mean kappa 0.54 versus 0.37).

Diagnostic accuracy

Diagnosis was more accurate with both FDG-PET methods than for any of the purely clinical methods (Table 5). Overall accuracy was superior ($P=0.02$), as was specificity for FTD ($P=0.02$). Diagnostic accuracy was consistently better using SSP than transaxial FDG-PET images, with 89% of all ratings having the correct diagnosis with SSP, but this did not reach statistical significance ($P=0.2$). The completion of a diagnostic checklist did not improve diagnostic accuracy. Diagnostic accuracy varied most among raters with the symptom checklist and least with clinical scenarios and SSP ratings. Diagnostic accuracy was less in FTD than in AD subjects (Figs 2–4). This discrepancy was present irrespective of the method used: clinical scenarios ($P=0.03$), transaxial images ($P=0.0003$) or SSP ($P=0.004$).

The usual methods for determining the value of a diagnostic test showed that imaging generally outperformed clinical measures. SSP achieved the highest sensitivities, specificities, predictive values and likelihood ratios of all four methods with the single exception of negative predictive value for FTD/positive predictive value for AD, where transaxial displays were slightly preferred (Tables 5 and 6). Of the clinical measures, the symptom checklist did not have an appreciable effect on accuracy or reliability. It provided a somewhat higher specificity for FTD, but it had a lower specificity for AD. It improved the positive likelihood ratio for FTD only slightly.

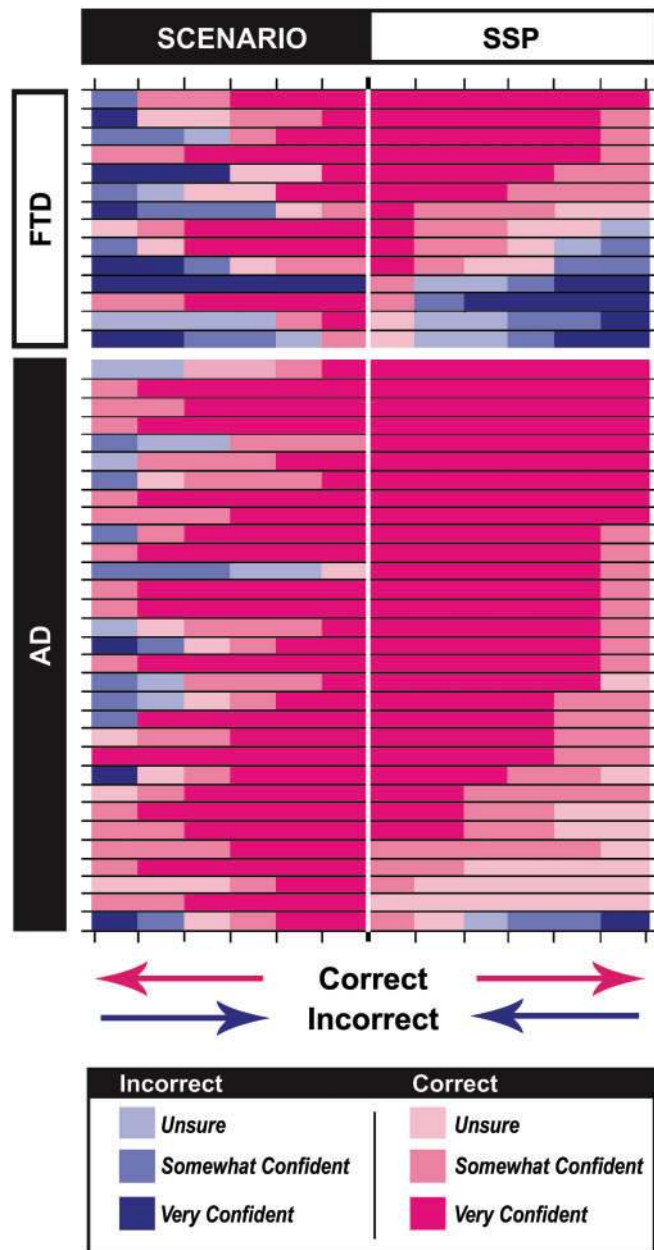


Fig. 2 Diagnostic accuracy and confidence after review of case scenario, and after review of FDG-PET scan displayed as stereotactic surface projection maps. Each horizontal bar represents the ratings in a single case. The length of the colour bar is proportionate to the number of ratings with a particular response. Vertical ticks represent the numbers of ratings, but not a particular individual rater. Cases are ordered vertically based upon the degree of diagnostic accuracy and confidence with SSP. Ratings with correct diagnoses are ordered with increasing confidence from the centre of the figure, while ratings with incorrect diagnoses are arranged from the outside with decreasing confidence. Ratings performed after review of the case scenario are shown in the left half of the figure, and those performed after review of the SSP FDG-PET scan are shown in the right half of the figure. Results in subjects with a neuropathological diagnosis of FTD are shown in the top of the figure, and of AD in the bottom of the figure. Clinical diagnoses judged to be correct as compared to pathological findings are shown in shades of red and incorrect diagnoses are

Diagnostic confidence

Raters often had limited confidence in their diagnosis, particularly with clinical methods (Figs. 2 and 3). Although completing a symptom checklist rarely altered a rater's diagnosis, it did tend to increase the rater's diagnostic confidence (data not shown). Raters might have inferred a degree of confidence based upon the value of the checklist score, but this was not the intent of the checklist authors or our instruction. Consequently, we did not evaluate confidence from the checklist alone. Diagnostic confidence appears to be a meaningful measure, because it appropriately reflected raters' true diagnostic accuracy. Raters were less confident about diagnosis when using transaxial images than SSP images ($P=0.02$) and less accurate about diagnosis when using scenarios than SSP images ($P=0.02$). They were also both less confident and less accurate in FTD cases than in AD cases. FDG-PET with SSP tended to improve the raters' diagnostic confidence more in AD than in FTD ($P=0.08$).

Rater performance

Overall the performance of raters was remarkably similar. Consistent with their status as dementia experts, the raters had similarly high diagnostic accuracy and all pair-wise inter-rater reliability comparisons were similar. All raters had better diagnostic accuracy with imaging methods than with the clinical scenario. All were more accurate with SSP images than transaxial images (Fig. 4). Raters all found it more difficult to accurately diagnose FTD than AD. The reliability and accuracy of ratings was unaffected by the rater's institutional affiliation, so we did not detect any geographical or practice variations.

Diagnostic accuracy of sequential ratings

The overall accuracy of initial clinical diagnosis of 79% improved to 90% ($P=0.03$) after FDG-PET scans were considered (Table 7). The addition of FDG-PET particularly improved the diagnostic accuracy of FTD ($P=0.01$), but improvements in the accuracy of AD did not reach statistical significance ($P=0.3$). Positive predictive value and positive likelihood ratio demonstrate that FDG-PET improves the clinical accuracy of an FTD diagnosis more than an AD diagnosis. Accuracy of an initial AD diagnosis after clinical information alone was 85%, leaving relatively

shown in shades of blue. The shading indicates the degree of certainty with most intense shades indicating greater degree of certainty. For example, the first row is an FTD case. All six raters were highly confident in a correct diagnosis of FTD after review of the SSP image. In contrast, after review of the clinical scenario, five raters indicated the correct diagnosis of FTD, three were highly confident and two were only somewhat confident. The sixth rater was somewhat confident in what ultimately was an incorrect diagnosis of AD. Overall diagnostic accuracy ($P=0.02$) and specificity for FTD ($P=0.02$) were significantly better after review of the SSP PET than after review of the case scenario.

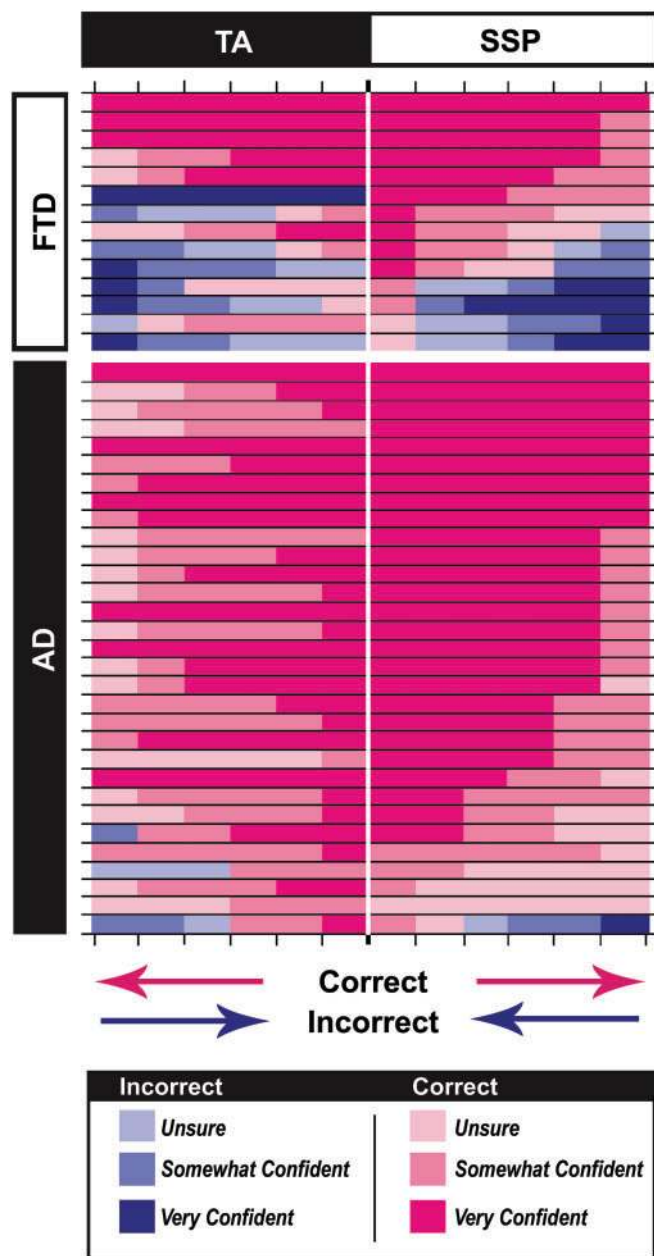


Fig. 3 Diagnostic accuracy and confidence with review of PET scans displayed as transaxial images, and after review of FDG-PET scans displayed as SSP images. The figure is constructed in the same way as in Fig. 2. Cases are ordered vertically based upon the degree of diagnostic accuracy and confidence with SSP. Although overall diagnostic accuracy were similar with the two methods, raters had significantly more confidence with SSP (difference in percentage very confident: overall, $P=0.02$, AD only $P=0.006$, FTD only 0.6; when restricted to cases where both diagnoses were correct: overall $P=0.006$, AD only $=0.005$, FTD only $P=0.8$).

little room for improvement. Regardless, the addition of FDG-PET increased the sensitivity for AD to 98%, and for 4 of 6 raters, to 100%.

Scans changed the diagnosis in 42 (16%) of the 270 ratings; 34 (81%) of which corrected an

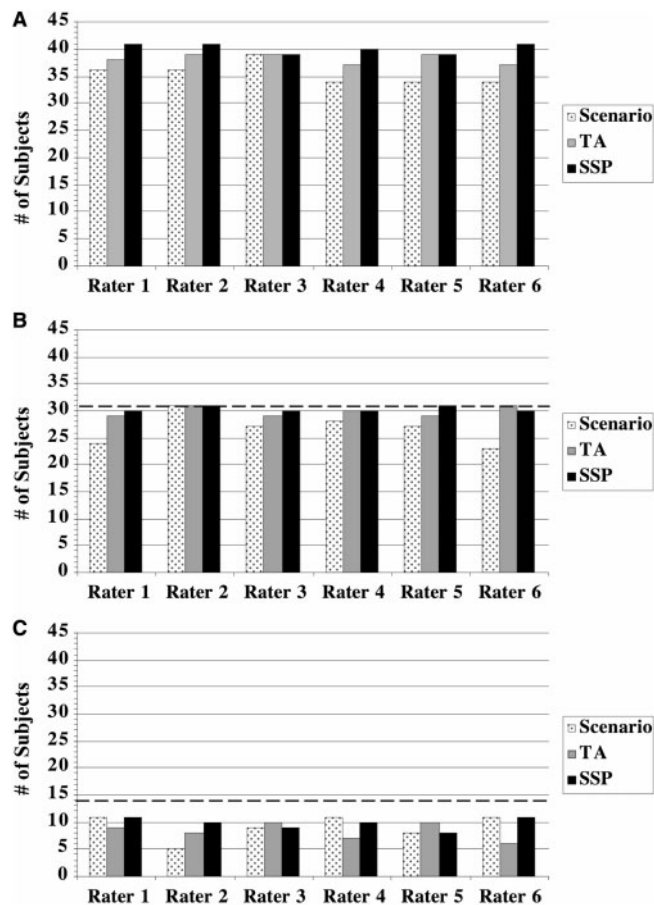


Fig. 4 Diagnostic accuracy by rater using three diagnostic methods—scenario alone, transaxial image alone and SSP image alone. In **A**, ratings of all 45 subjects are shown. In **B**, only the 31 AD subjects are shown and in **C**, only the 14 FTD subjects are shown. The y axis indicates the number of subjects. The same scale is used in all graphs with the dotted line showing the total possible number of subjects in **B** and **C**.

initial misdiagnosis. In 11 of these ratings, the rater changed their initial diagnosis of AD to FTD after review of the scan, which corrected a misdiagnosis in 10/11 (91%) cases. In 31 ratings, scan results led raters to change their diagnosis from FTD to AD, which corrected a misdiagnosis in 24/31 (77%) cases. The addition of FDG-PET caused diagnostic errors in 8/270 (3%) ratings: one error changed an initial AD diagnosis to FTD, and seven incorrectly changed initial FTD diagnoses to AD.

The frequency at which individual raters changed their diagnosis after viewing the FDG-PET scan was similar and ranged from 13 to 20%. The likelihood that a rater would change a diagnosis based upon FDG-PET scan results was not related to the extent of previous imaging experience or institutional affiliation.

Diagnostic confidence in sequential ratings

Viewing the FDG-PET scan was significantly more likely to increase diagnostic confidence than decrease it, even when

Table 6 Predictive values and likelihood ratios of five independent diagnostic methods

	Clinical scenario	Symptom checklist	Scenario + checklist	Transaxial FDG-PET	SSP FDG-PET
Mean PPV for FTD/NPV for AD (range of individual raters)	72% (56–100)	84% (58–100)	74% (56–100)	68% (50–88)	93% (85–100)
Mean NPV for FTD/PPV for AD (range of individual raters)	84% (78–90)	78% (71–89)	84% (78–90)	91% (87–97)	89% (85–92)
+ Likelihood ratio for FTD	4.5 (2.7–∞)	6.2 (3.0–∞)	4.6 (2.8–∞)	14.8 (10.7–∞)	36.5 (21.3–∞)
+ Likelihood ratio for AD	2.7 (1.6–4.3)	2.0 (1.1–3.5)	3.3 (1.6–4.4)	2.5 (1.8–3.2)	3.5 (2.3–4.6)
– Likelihood ratio for FTD	0.4 (0.2–0.6)	0.7 (0.3–0.9)	0.3 (0.2–0.6)	0.4 (0.3–0.6)	0.3 (0.2–0.4)
– Likelihood ratio for AD	0.2 (0–0.4)	0.2 (0–0.3)	0.2 (0–0.4)	0.2 (0–0.3)	0.03 (0–0.5)

Table 7 Accuracy of rater diagnoses before and after considering FDG-PET scans

	Before	After	P-value
Diagnostic accuracy, all cases (<i>n</i> = 45)	79% (73–82)	90% (78–98)	0.03
FTD sensitivity/AD specificity	65% (42–79)	71% (29–93)	0.6
FTD specificity/AD sensitivity	85% (71–97)	98% (90–100)	0.006
Positive predictive value for FTD/negative predictive value for AD	66% (71–97)	92% (71–100)	0.01
Negative predictive value for FTD/positive predictive value for AD	84% (79–88)	87% (76–94)	0.3
Positive likelihood ratio for FTD/negative likelihood ratio for AD ^a	4.3 (2.7–13.3)	44.3 (7.4–∞)	0.01
Positive likelihood ratio for AD/negative likelihood ratio for FTD ^a	2.4 (1.7–3.3)	3.4 (1.4–14.0)	0.5

Note: Values are the mean of all six raters. The range of individual raters is shown in parentheses.

^aA likelihood ratio of 1 indicates a test result does not alter pretest probability and has no value. A ratio of 2–5 indicates a small, but sometimes important, test. A ratio >10 is generally considered conclusive evidence that test performance changes pretest probability (Quizilbash, 2002).

diagnosis was unchanged ($P=0.003$, Table 8). Confidence appropriately reflected diagnostic accuracy (Table 8, Fig. 5). FDG-PET had most benefit on diagnostic accuracy when raters were only somewhat confident or uncertain about their initial diagnosis (Table 9). In these instances, diagnostic accuracy increased from 71 to 90% ($P=0.02$).

Overall effect of the addition of FDG-PET

The addition of FDG-PET changed a diagnosis or changed diagnostic confidence without changing the diagnosis in 53.7% of ratings (Fig. 6). We considered changes beneficial if they corrected a misdiagnosis, increased confidence in a

Table 8 Diagnostic confidence before and after considering FDG-PET scans^a

Diagnosis	Diagnostic confidence	Before (%)	After (%)
AD	Very confident	57	75
	Somewhat confident	34	20
	Uncertain	9	5
FTD	Very confident	30	65
	Somewhat confident	42	25
	Uncertain	28	10

^aP-values for the difference in the percentage of raters who were very confident: overall = 0.01, AD only = 0.01, FTD only = 0.02; when analysis was restricted to cases where diagnoses both before and after FDG-PET were correct, overall = 0.001, AD only = 0.03, FTD only = 0.0004.

correct diagnosis or decreased confidence in an incorrect diagnosis. Other changes were considered adverse. Using this guideline, the overall effect of adding FDG-PET on ratings was beneficial in 42.2%, neutral in 46.3% and adverse in 11.5%, and significantly more likely to be beneficial (42.2%) than adverse ($P=0.0001$).

Discussion

This study shows that FDG-PET is a reliable and valid diagnostic test that can aid physicians in making the sometimes difficult clinical distinction between AD and FTD. We believe brain imaging is most helpful in answering specific, narrowly framed, clinically relevant diagnostic questions. Consequently, this study was designed to evaluate only whether FDG-PET helped distinguish AD and FTD, which characteristically have sharply contrasting patterns of glucose hypometabolism. Our findings should encourage subsequent studies addressing how imaging practically can assist in answering the many other difficult diagnostic and management questions that arise during dementia evaluations.

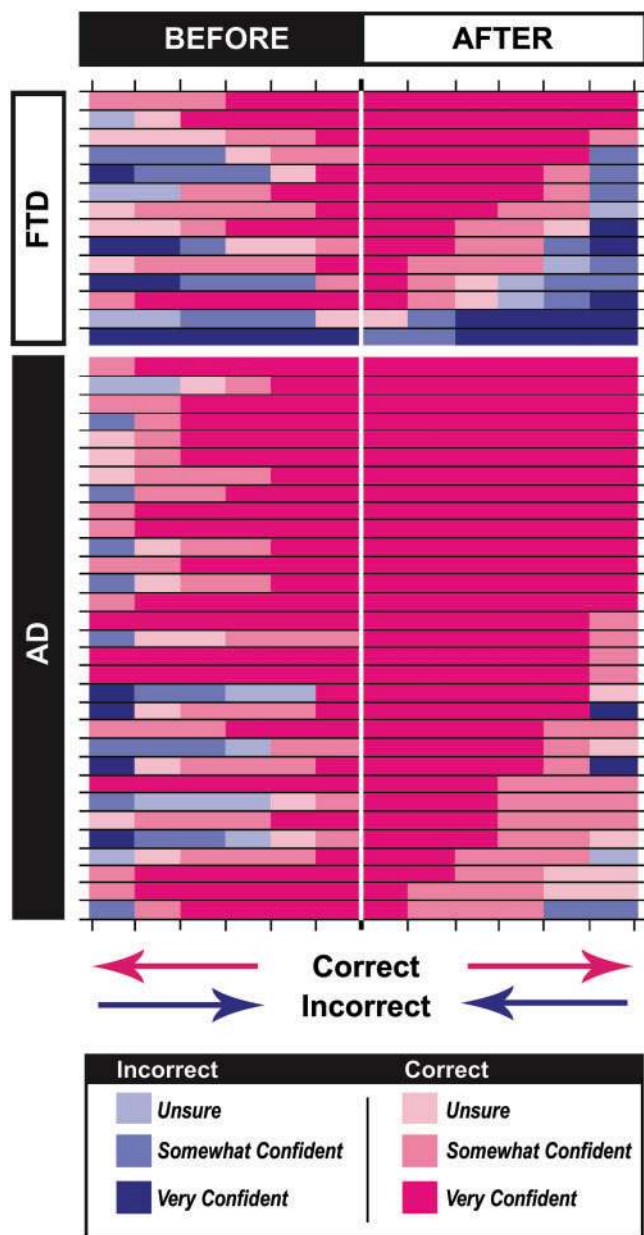


Fig. 5 Diagnosis and diagnostic confidence before and after review of FDG-PET scans. The figure is constructed in the same way as in Fig. 2. Ratings with only a review of history and examination are shown on the left of the figure, and ratings with FDG-PET added are on the right. Cases are ordered based upon the degree of accuracy and confidence found after FDG-PET was added to the clinical scenario. Overall diagnostic accuracy ($P=0.03$) and specificity for FTD ($P=0.006$) significantly improved when FDG-PET was added. Raters also had significantly greater diagnostic confidence (difference in percentage very confident: overall, $P=0.01$, AD only, $P=0.01$, FTD only $P=0.02$; restricted to cases where both diagnoses were correct: overall, $P=0.001$, AD only, $P=0.03$, FTD only $P=0.0004$).

Many previous FDG-PET studies have demonstrated that these two disorders cause distinctive patterns of glucose hypometabolism (Ishii *et al.*, 1998; Foster *et al.*, 1999; Foster, 2003b). These and most studies of FDG-PET have

Table 9 Effect of initial diagnostic confidence on changes in diagnosis and accuracy

	Diagnosis correct before FDG-PET	Diagnosis changed after FDG-PET	Diagnosis correct after FDG-PET
Very confident (132 ratings)	115 (87%)	6 (5%)	115 (87%)
Somewhat confident or uncertain (138 ratings)	98 (71%)	36 (26%)	124 (90%)

focused on group differences. Voxel-wise analysis methods also have identified changes in groups of subjects with mild cognitive impairment before the development of frank dementia (Chetelat *et al.*, 2003; Anchisi *et al.*, 2005). More stringent requirements must be met to demonstrate that these differences are applicable to individuals and can be used in the practical care of patients. Several studies are available showing that dementing diseases cause metabolic changes that are sufficiently robust to be identified with FDG-PET in single subjects (Salmon, 2002). One large, multi-centre study of FDG-PET used individual differences from a reference population to distinguish individual normal control subjects and patients with clinically diagnosed AD (Herholz *et al.*, 2002). Using an automated analysis method to quantify the sum of *t*-score values across all abnormal voxels, the sensitivity and specificity to discriminate between mild-moderate AD and normal subjects were both 93%. Diagnostic classification, however, was based upon a *post hoc* criterion derived from the study data. Considerable evidence also has accumulated demonstrating that visual interpretation of FDG-PET has a high degree of diagnostic accuracy when compared to neuropathological diagnosis (Mielke *et al.*, 1996; Hoffman *et al.*, 2000; Silverman *et al.*, 2001). However, in these studies no comparison with clinical information was possible. This study expands on previous studies by showing that visual interpretation of FDG-PET with predetermined diagnostic criteria can have high inter-rater reliability, is superior to a detailed clinical summary, and when added to other clinical information, can enhance the accuracy and confidence of diagnosis.

Implications for clinical assessment

Imaging is unreliable as the sole basis of determining the cause of dementia and only should be used as an adjunct to other components of the diagnostic evaluation. A careful consideration of the medical history and examination will continue to be essential to dementia evaluations. However, the recognition and interpretation of symptoms and disease course are subjective, and the accuracy and confidence of diagnosis using clinical methods alone can vary, depending upon patient and physician characteristics and the amount

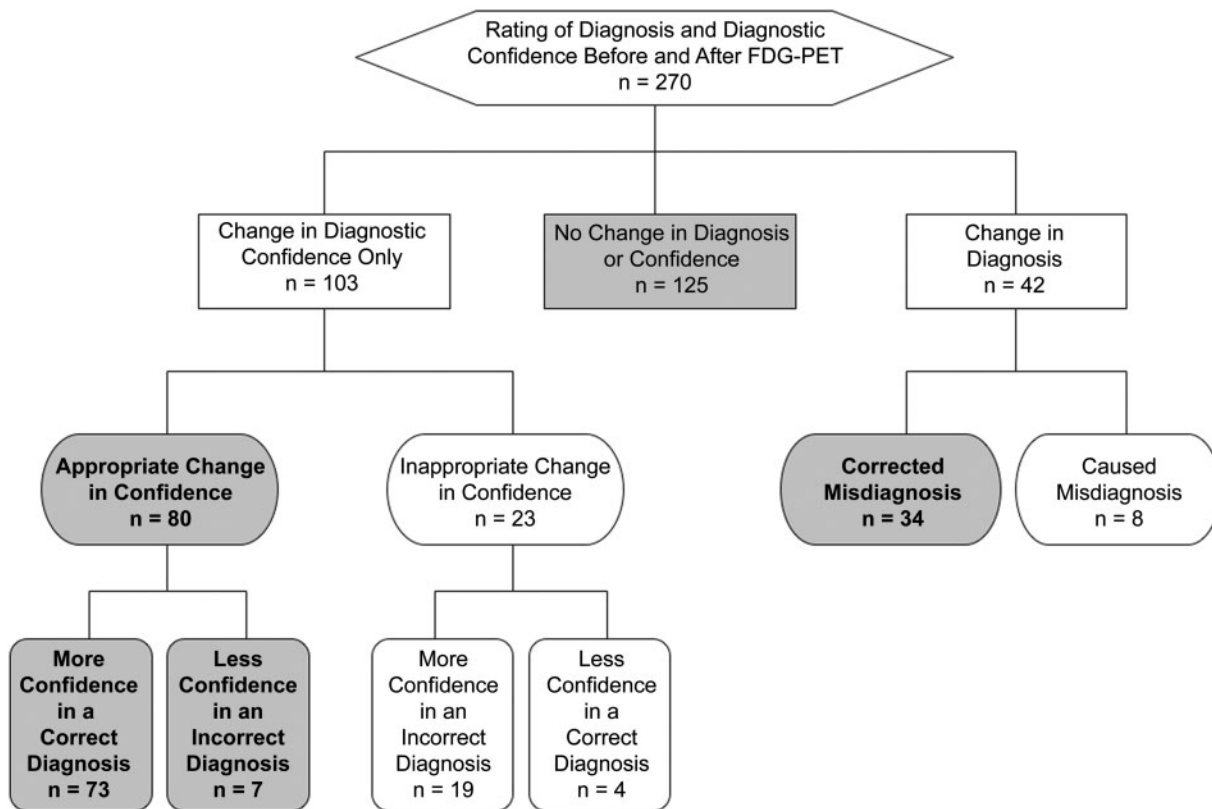


Fig. 6 The effect of FDG-PET on diagnostic accuracy and confidence. This flow diagram shows the overall effect of FDG-PET studies on diagnosis and diagnostic confidence. Outcomes that are beneficial are indicated in bold text. Outcomes that are either neutral or beneficial are shaded. Appropriate changes in diagnosis after adding FDG-PET were significantly greater than inappropriate changes ($P = 0.0001$).

and quality of information available. As our data illustrate, even experts find it difficult to diagnose a specific dementing disease in some cases. We asked our raters to rely upon their clinical expertise rather than use explicit diagnostic criteria. We chose this approach because it more accurately reflected clinical practice, we were separately evaluating a diagnostic checklist, and proposed diagnostic criteria for FTD are still being refined and have not yet had neuropathological validation (Neary *et al.*, 1998; Rosen *et al.*, 2002). Furthermore, it was not our intention to evaluate diagnostic criteria or propose new operational procedures for criteria that often are subject to considerable interpretation. Relying on our expert judgements appears justified because there generally was high agreement about diagnoses and we were unable to identify any significant inter-rater or institutional differences. Furthermore, raters achieved a sensitivity and specificity for AD based upon the clinical scenarios well within the range observed in other studies (Chui and Lee, 2002). While fewer studies of FTD are available, they also achieved a sensitivity and specificity for FTD based upon clinical scenario that is consistent with other retrospective studies. For example, one study of eight FTD patients found a mean sensitivity of 85%, a mean specificity of 97% and a mean kappa value of 0.75 using only first visit clinical data and Lund and Manchester clinical criteria (Lopez *et al.*, 1999). Another study of seven

patients with Pick's disease that did not use clinical criteria found a median sensitivity of 43%, a median specificity of 99% and a mean kappa value of 0.42 using only first visit information (Litvan *et al.*, 1997).

It has been argued that FDG-PET can add little to dementia evaluations because of high accuracy in the diagnosis of AD, at least in research centers studying Alzheimer's disease and when experts perform the evaluations (Lopez *et al.*, 1999; Holmes *et al.*, 1999; Chui and Lee, 2002; Gill *et al.*, 2003). However, this partly reflects the high prior probability of AD in most clinical populations, and does not necessarily imply a similar accuracy in clinical situations where AD is less likely. In our study, adding FDG-PET to clinical information increased the accuracy of AD diagnosis from 86 to 97%. Diagnostic accuracy of FTD using only clinical information is less (Litvan *et al.*, 1997). Recognizing the challenges of differentiating FTD and AD using only clinical history and examination, several diagnostic checklists have been proposed, each intended to focus clinicians' attention to symptoms felt to be characteristic of FTD (Miller *et al.*, 1997; Kertesz *et al.*, 2000). We chose to use the checklist of Barber *et al.*, because it identifies symptoms characteristic of FTD and AD to generate a score for diagnosis, was designed for retrospective review of a patient's entire clinical course, and is the only checklist that has been validated

in autopsy-confirmed cases (Barber *et al.*, 1995). Unfortunately, we found the checklist had poor reliability. While the Barber checklist was intended for use with reports of knowledgeable informants rather than review of medical records, this does not seem to explain its weaknesses in our study. Completing the diagnostic checklist during review of the scenario did not improve the diagnostic accuracy of our expert raters, although it might benefit less experienced clinicians.

Interpretation of FDG-PET scans

Both diagnostic checklists and usual clinical methods distinguish AD and FTD primarily by inferring sites of pathology based on characteristic signs and symptoms. FDG-PET provides a separate, simpler, more objective and quantitative way to make this same judgement. Imaging provides a wealth of data, which can be displayed in many ways. Good reliability and diagnostic accuracy were observed with traditional transaxial images, but even better results were achieved when SSP maps were used to display FDG-PET data. The advantage of SSP presumably is due to its ability to summarize data from the cerebral cortex in a few images and provide a visual comparison with findings in normal control subjects. We found that SSP was inferior to transaxial images only in its lesser reliability in judging ‘significant’ hemispheric asymmetry, a decision not aided by current SSP methods. Statistical maps that highlight hemispheric differences likely would remedy this weakness. There also are many potential ways to analyse FDG-PET data. It is reassuring that we achieved substantial levels of inter-rater reliability after a relatively brief training using a simple diagnostic rule with the clinically practical approach of visual interpretation. Visual assessment of images has the advantage of utilizing all available information and permits clinicians to simultaneously consider many factors that may have diagnostic significance. Indeed, we have found in a separate study using these same subjects that visual interpretation of scans for diagnosis is superior to automated algorithms comparing metabolic changes in specific brain regions (Higdon *et al.*, 2004). On the other hand, multivariate predictive models using partial least-squares analysis based on the raw data from the SSP images were able to achieve diagnostic accuracy similar to that of our expert raters. Raters found that some scans did not have clear-cut abnormalities, particularly those with very mild deficits. Future studies should evaluate strategies that could increase sensitivity in detecting abnormalities in individual scans such as multivariate statistical methods and using cohort-specific normative brain atlases for statistical analyses. The lower inter-rater reliability in judging abnormality in individual brain regions suggests that reliability would be poor for an algorithm derived from a combination of regions. Likewise, visual assessment of metabolic asymmetry appears to be insufficiently reliable as a basis for diagnostic decisions.

Diagnostic confidence

Our results demonstrate that FDG-PET increases diagnostic confidence. Appropriately enhancing diagnostic confidence could have a major benefit for patients. The uncertainty that physicians feel in diagnosing dementing diseases has been little studied, but likely is an important factor in the quality of dementia care (Foster, 2001). Presumably, physicians more confident in their diagnoses are more likely to institute and sustain therapy, disclose a diagnosis and fully discuss prognosis with patients and families. It is difficult to know how diagnostic uncertainty in this study might compare to the diagnostic confidence physicians experience in clinical practice. Undoubtedly, individual style and personality must heavily influence confidence. We found that even highly experienced dementia experts have considerable and remarkably similar degrees of uncertainty in their diagnosis when it is based solely on clinical data. It is noteworthy that our expert raters all found the diagnosis obvious in only a few cases; in the vast majority of cases, different raters expressed a range of diagnostic confidence when provided with only clinical information. We suspect that community physicians with less experience in the diagnosis of dementing diseases would express even greater uncertainty. This partly may account for the significant underutilization of available treatments in the community, even when dementia is recognized (Magsi *et al.*, 2005). Our study suggests that diagnostic confidence is a valid reflection of diagnostic accuracy. It warrants further study and should be considered as an outcome to assess the value of a diagnostic test, because diagnostic confidence is likely to affect physicians’ decisions about treatment. Rater uncertainty varied considerably from case to case, and not surprisingly, was greater for patients with FTD. FTD has many diverse presentations that can be subtle or overlap with AD, and since it is less common, clinicians also have less experience to draw upon.

Potential limitations

It is crucial to determine the cause of dementia early so that treatments can begin. This study only included patients with autopsy-confirmed diagnoses so we could use a generally accepted gold standard for judging diagnostic accuracy. Initial clinical evaluations were performed without the benefit of our current understanding of dementing diseases and often many years after symptom onset. Our study provides critical information about histopathological diagnosis, which is difficult to obtain in a substantial cohort of subjects who were scanned with very mild or pre-clinical dementia. At least in the period represented in this study, dementia evaluations were often delayed for many years after symptom onset and few were scanned with very mild or pre-clinical dementia.

The retrospective review of clinical summaries is artificial and autopsy confirmation required in this study

mean that the subjects were necessarily highly selected and do not reflect a diverse population required for a Class I study. Consequently, our results, including specificities and sensitivities, may not be replicated in clinical practice. Although a dementia expert extracted clinical information most relevant to diagnostic decision-making, case scenarios do not replicate the usual interactions between a physician, patient and knowledgeable informants. On the other hand, it is difficult to know whether our procedures would be more or less likely to provide an accurate diagnosis than usual clinical practice. Our use of dementia experts likely increased the accuracy of diagnoses that were based solely on clinical data. Clinical diagnosis is more often confirmed at autopsy when specialists rather than community physicians perform dementia evaluations (Becker *et al.*, 1994; Holmes *et al.*, 1999; Mendez *et al.*, 1992). Our summaries were derived from extensive medical records at dementia research centres, rather than less extensive assessments performed in the community. Furthermore, our case summaries described the entire course of a patient's illness, information unavailable at an initial diagnostic assessment. Although it is conceivable that knowledge of symptoms occurring later in the illness could be confusing and detrimental, previous studies have shown that diagnostic accuracy improves when longitudinal information is available (Becker *et al.*, 1994; Litvan *et al.*, 1997). It also is important to recognize that patients in this study may differ from those encountered in typical clinical practice. Although some of our subjects were only mildly impaired and others were severely demented when they were scanned, the severity of their dementia in our study population may not be representative of patients presenting a similar diagnostic dilemma. We used the sequential review of a clinical history and examination followed by a review of imaging results to better reflect good medical practice. However, a prospective trial of FDG-PET in the evaluation of suspected FTD would complement this study and could adequately address many of its limitations.

Our study forced raters to make a diagnostic decision, even if they felt the FDG-PET scan was normal. The raters were aware that all of the cases presented had dementia and a diagnosis of 'normal' was not allowed. This encouraged the raters to consider even minor variations in the scan. Although it is not typical to force radiologists to make a diagnosis without characteristic image findings, clinicians often need to make definitive decisions, even when confronted by an ambiguous history and examination. Thus it seemed only fair to force the raters to always make a specific diagnosis. It is unclear whether this would cause any systematic bias, but the results appear to justify the approach.

Alternative approaches

Several other methods also may help physicians distinguish AD from FTD. We did not include structural brain imaging

in our study, but it is possible that CT or MRI also would aid diagnostic accuracy. CT and MRI scans can reveal regional atrophy, which might aid in diagnosis (Chan *et al.*, 2001; Boccardi *et al.*, 2003). However, visual assessment of hippocampal atrophy is not helpful in distinguishing AD and FTD (Galton *et al.*, 2001). We also did not explicitly consider results of neuropsychological testing, although pertinent findings were recorded in the summaries for many of our cases. The possible contribution to diagnosis of standardized psychological testing should be evaluated. Many body fluid and other imaging biomarkers for AD and FTD have been proposed and are under investigation (Frank *et al.*, 2003; Petrella *et al.*, 2003; Klunk *et al.*, 2004). They may prove valuable in the future, but require systematic validation following consensus guidelines (The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group, 1998).

FDG-PET as a diagnostic biomarker

It is appropriate to consider the status of FDG-PET as a diagnostic biomarker in light of the current results. FDG-PET is clearly imperfect. Although raters knew that all cases had an autopsy-confirmed dementing illness, 16% of the transaxial images and 5% of the SSP images were rated as normal or having uncertain abnormality and likely would be considered non-diagnostic in a clinical setting. The proportion of ratings indicating that scans did not have clear-cut abnormalities was similar for both FTD and AD subjects. On the other hand, this study substantially adds to the increasing evidence that FDG-PET has utility as a diagnostic biomarker as judged by guidelines recommended by the Reagan Institute–NIA Work Group. FDG-PET meets many of the proposed ideal characteristics. It reflects a fundamental pathological feature of AD and FTD, the selective regional loss of neurons and synapses in the cerebral cortex. It has been validated using neuropathologically confirmed cases in this and other studies (Hoffman *et al.*, 2000; Silverman *et al.*, 2001). It also is notable that in this study we have shown that FDG-PET is reliable and can distinguish between two pathologically distinct causes of dementia. Most studies of diagnostic biomarkers only compare AD with normal subjects or with non-demented controls, a much less clinically relevant distinction. Our study included several individuals who were scanned when they had mild dementia. Our results confirm the observations of others that FDG-PET is able to detect disease early in its course, as the biomarker guidelines recommend (Minoshima *et al.*, 1997; Berent *et al.*, 1999; Chetelat *et al.*, 2003). Our study does not address the remaining ideal characteristics of a diagnostic biomarker. It will remain a subjective opinion whether FDG-PET is sufficiently non-invasive, simple to perform and inexpensive to meet these guideline criteria. Widespread experience over the past several years has

demonstrated that FDG-PET can be practical in clinical diagnosis, even if it is not ideal by these standards.

Most of the steps recommended by the Reagan Institute–NIA Work Group for establishing FDG-PET as a biomarker have been achieved. In our study, the traditional standards of sensitivity, specificity and positive predictive value have been achieved for FDG-PET using SSP, except for FTD sensitivity and AD specificity, which are 73% rather than 80%. Positive likelihood ratio (+LR), a measure of how a test alters pretest probability also should be considered (Qizilbash, 2002). Tests with +LR values >10 generally are considered to make conclusive changes to pretest probability. By this standard, our study finds FDG-PET has great value for patients with FTD, but the pre-test probability of AD is too high for a test to have much effect on the +LR, even if it is highly accurate. There are other ways to assess the value of FDG-PET, including its effect on diagnostic confidence, treatment decisions and health costs (Gill *et al.*, 2003; McMahon *et al.*, 2003).

Practical implications

Great care is needed when incorporating FDG-PET into the diagnostic evaluation of patients with dementia. Imaging cannot substitute for clinical information including a detailed history, neurological and mental status examinations. FDG-PET scan results cannot be considered conclusive and there is a risk of misinterpretation of scan results. Pathological assessment still remains the only certain way to confirm a clinical diagnosis. Physicians will continue to face complexity when trying to distinguish AD from FTD as illustrated by several recent examples in the literature. A single case has been reported with frontal predominant clinical features and AD pathology that appeared to be more severe in the frontal cortex (Johnson *et al.*, 1999). This raises the possibility that AD might sometimes cause frontal predominant hypometabolism. In another report, a patient with clinical symptoms of frontotemporal dementia was initially thought to have AD because of a presenilin-1 gene mutation (PSEN1ins352). Subsequently, this patient was found to have both FTD ubiquitin inclusion pathology and a splice donor site mutation in intron 1 of the progranulin gene. The presenilin gene mutation in this case now appears to be a non-pathological variant (Pickering-Brown *et al.*, 2006). In this case, clinical symptoms were more reliable than a genetic test. Unfortunately, molecular imaging results are unavailable for either of these cases, so we do not know whether FDG-PET could have aided accurate diagnosis. However, the report of an Italian family supports our findings and found FDG-PET more diagnostically reliable than clinical symptoms (Binetti *et al.*, 2003). In affected members of this family clinical history and symptom presentation suggested a frontotemporal dementia and led to an unsuccessful search for a mutation in the TAU gene. Instead, FDG-PET revealed a clear-cut AD pattern of

hypometabolism and a novel T122R mutation in the presenilin-2 gene (PSEN2) was subsequently discovered.

Our study addresses only one of many diagnostic decision points physicians face in determining the cause of dementia. FDG-PET scans may or may not be similarly helpful when other disorders are suspected in addition to FTD and AD. For example, in patients suspected to have FTD, psychiatric illness often is a consideration and has not been addressed in our study. Our subjects did not have other neurological disorders and therefore a diagnosis reasonably could be made considering only the FDG-PET scan. In clinical practice, when complicating brain diseases may be present, FDG-PET scans should only be interpreted in conjunction with structural imaging studies. We decided to compare the test characteristics of transaxial and SSP imaging because of previous evidence that SSP was a superior method for dementia diagnosis (Burdette *et al.*, 1996). This meant that raters were not able to directly compare the two methods. Combining the two methods may well have advantages. However, more information is not always better for diagnostic interpretation and summarizing imaging data may have its merits. Current PET scanners often provide 128 or more transaxial images, which can be challenging to mentally combine and manipulate. Further research evaluating image interpretation is needed. Although we chose SSP because of our familiarity and theoretical advantages, there are several other software packages available that display images topographically and provide statistical maps and may provide similar benefits.

Training of raters was an important factor in achieving the reliability and accuracy of FDG-PET interpretation in this study. The experience of the physicians interpreting clinical scans should be considered when ordering FDG-PET studies. Since clinical use is a relatively recent development, few physicians have been trained to evaluate FDG-PET scans for the complex pattern of hypometabolism seen in patients with dementia. Fortunately, our study found that relatively brief, focused training given to clinicians experienced in the evaluation of dementia is adequate to provide good diagnostic precision in the interpretation of FDG-PET scans. It is critical to attend to technical issues involved with image acquisition and processing. Errors in attenuation correction cause normal scans to have the appearance of AD or FTD. Patient behaviour during scanning also can lead to misinterpretation of results. Our subjects were all scanned at rest with eyes open in a quiet, darkened room. Our rules for scan interpretation may not apply to other protocols.

Diagnostic accuracy was enhanced most when the clinician had limited diagnostic confidence after considering only the results of a clinical history and examination. This suggests that a targeted approach using FDG-PET when there is diagnostic uncertainty would have greatest clinical value. We found diagnostic confidence was consistently less in patients with FTD and FDG-PET often

improved accuracy and confidence. In some circumstances, diagnostic confidence also may be low in some patients with AD, particularly in those who have atypical presentations. Variations in the clinical presentation of AD are being recognized increasingly (Caselli, 2000; Nestor *et al.*, 2003; Tang-Wai *et al.*, 2004; Knibb *et al.*, 2006). Enhancing diagnostic accuracy and confidence in these situations will have a favourable impact on patient management.

Summary

FDG-PET is a promising diagnostic biomarker in dementing illness. We have shown that it is valuable in differentiating AD and FTD, and particularly helpful when findings in a clinical evaluation are not definitive and physicians are not already highly confident in their diagnosis. While additional research is needed to assure that FDG-PET is used to greatest advantage in dementia care, the visual interpretation of FDG-PET scans using voxel-based analysis with our simple diagnostic rules to distinguish AD and FTD provides a practical approach that can be widely applied now to enhance diagnosis.

Acknowledgements

This study was supported by a pilot cooperative project grant from the National Alzheimer's Coordinating Center (NIH Grant AG16976) and by the Michigan Alzheimer's Disease Research Center (NIH Grant AG08671), the University of California at Davis Alzheimer's Center (NIH Grant AG10129) and the University of Pennsylvania Alzheimer's Center (NIH Grant AG10124). Co-author Satoshi Minoshima developed and owns the copyright to the Neurostat Neurological Statistical Imaging Software used in this project. Dr Foster has received an honorarium of less than \$5000 for serving on the Scientific Advisory Board of GE Healthcare. We thank Andrew P. Lieberman for reviewing the neuropathology, and David E. Kuhl, Sid Gilman, Gus Buchtel, David Knesper, R. Scott Turner and Kirk Frey for making images from their research available for this study. Members of the panel who provided a retrospective consensus diagnosis for these subjects were Howard Aizenstein, Bradley Boeve, James Leverenz, Elaine Peskind, Kyle Womack and Edward Zamrini.

References

- Anchisi D, Borroni B, Franceschi M, Kerrouche N, Kalbe E, Beuthien-Beumann B, et al. Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease. *Arch Neurol* 2005; 62: 1728–33.
- Andrews DWK. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 1991; 59: 817–58.
- Barber R, Snowden JS, Craufurd D. Frontotemporal dementia and Alzheimer's disease: retrospective differentiation using information from informants. *J Neurol Neurosurg Psychiatry* 1995; 59: 61–70.
- Becker JT, Boller F, Lopez OL, Saxton J, McGonigle KL. The natural history of Alzheimer's disease. Description of study cohort and accuracy of diagnosis. *Arch Neurol* 1994; 51: 585–94.
- Binetti G, Signorini S, Squitti R, Alberici A, Benussi L, Cassetta E, et al. Atypical dementia associated with a novel presenilin-2 mutation. *Ann Neurol* 2003; 54: 832–6.
- Berent S, Giordani B, Foster N, Minoshima S, Lajiness-O'Neill R, Koeppel R, et al. Neuropsychological function and cerebral glucose utilization in isolated memory impairment and Alzheimer's disease. *J Psychiatr Res* 1999; 33: 7–16.
- Blacker D, Albert MS, Bassett SS, Go RC, Harrell LE, Folstein MF. Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative. *Arch Neurol* 1994; 51: 1198–204.
- Boccardi M, Laakso MP, Bresciani L, Galluzzi S, Geroldi C, Beltramello A, et al. The MRI pattern of frontal and temporal brain atrophy in fronto-temporal dementia. *Neurobiol Aging* 2003; 24: 95–103.
- Burdette JH, Minoshima S, Vander Borght T, Tran DD, Kuhl DE. Alzheimer disease: improved visual interpretation of PET images by using three-dimensional stereotaxic surface projections. *Radiology* 1996; 198: 837–43.
- Caselli RJ. Focal and asymmetric cortical degenerative syndromes. *Adv Neurol* 2000; 82: 35–51.
- Chan D, Fox NC, Jenkins R, Scallan RI, Crum WR, Rossor MN. Rates of global and regional cerebral atrophy in AD and frontotemporal dementia. *Neurology* 2001; 57: 1756–63.
- Chetelat G, Desgranges B, de la Sayette V, Viader F, Eustache F, Baron JC. Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? *Neurology* 2003; 60: 1374–77.
- Chui H, Lee A-Y. Clinical criteria for dementia subtypes. In: Qizilbash N, editor. Evidence-based dementia practice. Oxford: Blackwell Science; 2002. p. 106–19.
- Foster NL. Barriers to treatment: the unique challenges for physicians providing dementia care. *J Geriatr Psychiatry Neurol* 2001; 14: 188–98.
- Foster NL. Neuropsychiatry of dementing disorders. In: Fogel BS, Schiffer RB, Rao SM, editors. *Neuropsychiatry*. 2nd edn. New York: Lippincott Williams & Wilkins; 2003a. p. 1034–70.
- Foster NL. Validating FDG-PET as a biomarker for frontotemporal dementia. *Exp Neurol* 2003b; 184 (Suppl 1): S2–8.
- Foster NL, Minoshima S, Kuhl DE. Brain imaging in Alzheimer disease. In: Terry RD, Katzman R, Bick KL, Sisodia SS, editors. *Alzheimer Disease*. 2nd edn. New York: Raven Press; 1999. p. 67–93.
- Frank RA, Galasko D, Hampel H, Hardy J, de Leon MJ, Mehta PD, et al. Biological markers for therapeutic trials in Alzheimer's disease: proceedings of the Biological Markers Working Group, NIA Initiative on Neuroimaging in Alzheimer's Disease. *Neurobiol Aging* 2003; 24: 521–36.
- Galton CJ, Gomez-Anson B, Antoun N, Scheltens P, Patterson K, Graves M, et al. Temporal lobe rating scale: application to Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 2001; 70: 165–73.
- Gill SS, Rochon PA, Guttman M, Laupacis A. The value of positron emission tomography in the clinical evaluation of dementia. *J Am Geriatr Soc* 2003; 51: 258–64.
- Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 2002; 297: 353–56.
- Herholz K, Salmon E, Perani D, Baron JC, Holtoff V, Frolich L, et al. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage* 2002; 17: 302–16.
- Higdon R, Foster NL, Koeppel RA, DeCarli CS, Jagust WJ, Clark CM, et al. A comparison of classification methods for differentiating fronto-temporal

- dementia from Alzheimer's disease using FDG-PET imaging. *Stat Med* 2004; 23: 315–26.
- Hoffman JM, Welsh-Bohmer KA, Hanson M, Crain B, Hulette C, Earl N, et al. FDG PET imaging in patients with pathologically verified dementia. *J Nucl Med* 2000; 41: 1920–28.
- Holmes C, Cairns N, Lantos P, Mann A. Validity of current clinical criteria for Alzheimer's disease, vascular dementia and dementia with Lewy bodies. *Br J Psychiatry* 1999; 174: 45–50.
- Ishii K, Sakamoto S, Sasaki M, Kitagaki H, Yamaji S, Hashimoto M, et al. Cerebral glucose metabolism in patients with frontotemporal dementia. *J Nucl Med* 1998; 39: 1875–78.
- Ishii K, Willoch F, Minoshima S, Drzezga A, Ficarò EP, Cross DJ, et al. Statistical brain mapping of 18F-FDG PET in Alzheimer's disease: validation of anatomic standardization for atrophied brains. *J Nucl Med* 2001; 42: 548–57.
- Johnson JK, Head E, Kim R, Starr A, Cotman CW. Clinical and pathological evidence for a frontal variant of Alzheimer disease. *Arch Neurol* 1999; 56: 1233–9.
- Kertesz A, Munoz D. Pick's disease, frontotemporal dementia, and Pick complex: emerging concepts. *Arch Neurol* 1998; 55: 302–4.
- Kertesz A, Nadkarni N, Davidson W, Thomas AW. The frontal behavioral inventory in the differential diagnosis of frontotemporal dementia. *J Int Neuropsychol Soc* 2000; 6: 460–68.
- Klunk WE, Engler H, Nordberg A, Wang Y, Blomqvist G, Holt DP, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol* 2004; 55: 306–19.
- Knibb JA, Xuereb JH, Patterson K, Hodges JR. Clinical and pathological characterization of progressive aphasia. *Ann Neurol* 2006; 59: 156–65.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
- Litvan I, Agid Y, Sastry N, Jankovic J, Wenning GK, Goetz CG, et al. What are the obstacles for an accurate clinical diagnosis of Pick's disease? A clinicopathologic study. *Neurology* 1997; 49: 62–9.
- Lopez OL, Litvan I, Catt KE, Stowe R, Klunk W, Kaufer DJ, et al. Accuracy of four clinical diagnostic criteria for the diagnosis of neurodegenerative dementias. *Neurology* 1999; 53: 1292–99.
- Magsi H, Malloy T. Underrecognition of cognitive impairment in assisted living facilities. *J Am Geriatr Soc* 2005; 53: 295–98.
- McKeith IG, Galasko D, Kosaka K, Perry EK, Dickson DW, Hansen LA, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology* 1996; 47: 1113–24.
- McMahon PM, Araki SS, Sandberg EA, Neumann PJ, Gazelle GS. Cost-effectiveness of PET in the diagnosis of Alzheimer disease. *Radiology* 2003; 228: 515–22.
- Mendez MF, Mastri AR, Sung JH, Frey WH, 2nd. Clinically diagnosed Alzheimer disease: neuropathologic findings in 650 cases. *Alzheimer Dis Assoc Disorder* 1992; 6: 35–43.
- Mendez MF, Selwood A, Mastri AR, Frey WH, 2nd. Pick's disease versus Alzheimer's disease: a comparison of clinical characteristics. *Neurology* 1993; 43: 289–92.
- Mielke R, Schroder R, Fink GR, Kessler J, Herholz K, Heiss WD. Regional cerebral glucose metabolism and postmortem pathology in Alzheimer's disease. *Acta Neuropathol* 1996; 91: 174–79.
- Miller BL, Ikonte C, Ponton M, Levy M, Boone K, Darby A, et al. A study of the Lund-Manchester research criteria for frontotemporal dementia: clinical and single-photon emission CT correlations. *Neurology* 1997; 48: 937–42.
- Minoshima S, Frey KA, Foster NL, Kuhl DE. Preserved pontine glucose metabolism in Alzheimer's disease: a reference region for functional brain analysis. *J Comput Assist Tomogr* 1995a; 19: 541–47.
- Minoshima S, Frey KA, Koeppe RA, Foster NL, Kuhl DE. A diagnostic approach in Alzheimer's disease using 3-dimensional stereotactic surface projections of fluorine-18-FDG PET. *J Nucl Med* 1995b; 36: 1238–48.
- Minoshima S, Giordani BJ, Berent S, Frey KA, Foster NL, Kuhl DE. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neurol* 1997; 42: 85–94.
- Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 1998; 51: 1546–54.
- Nestor PJ, Caine D, Fryer TD, Clarke J, Hodges JR. The topography of metabolic deficits in posterior cortical atrophy (the visual variant of Alzheimer's disease) with FDG-PET. *J Neurol Neurosurg Psychiatry* 2003; 74: 1521–9.
- Petrella JR, Coleman RE, Doraiswamy PM. Neuroimaging and early diagnosis of Alzheimer disease: a look to the future. *Radiology* 2003; 226: 315–36.
- Pickering-Brown SM, Baker M, Gass J, Boeve BF, Loy CT, Brooks WS, et al. Mutations in progranulin explain atypical phenotypes with variants in MAPT. *Brain* 2006; 129: 3124–6.
- Procter AW, Qurne M, Francis PT. Neurochemical features of frontotemporal dementia. *Dement Geriatr Cogn Disord* 1999; 10: 80–4.
- Qizilbash N. Evidence-based diagnosis. In: Qizilbash N, editor. *Evidence-based dementia practice*. Oxford: Blackwell Science; 2002. p. 18–25.
- Ratnavalli E, Brayne C, Dawson K, Hodges JR. The prevalence of frontotemporal dementia. *Neurology* 2002; 58: 1615–21.
- Rosen HJ, Hartikainen KM, Jagust W, Kramer JH, Reed BR, Cummings JL, et al. Utility of clinical criteria in differentiating frontotemporal lobar degeneration (FTLD) from AD. *Neurology* 2002; 58: 1608–15.
- Rosso SM. Frontotemporal dementia in the Netherlands. Rotterdam: Optima Grafische Communicatie; 2003.
- Salmon E. Functional brain imaging applications to differential diagnosis in the dementias. *Curr Opin Neurol* 2002; 15: 439–44.
- Silverman DH, Small GW, Chang CY, Lu CS, Kung De Aburto MA, et al. Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome. *JAMA* 2001; 286: 2120–27.
- Tang-Wai DF, Graff-Radford NR, Boeve BF, Dickson DW, Parisi JE, Crook R, et al. Clinical, genetic, and neuropathologic characteristics of posterior cortical atrophy. *Neurology* 2004; 63: 1168–74.
- The National Institute on Aging and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. *Neurobiol Aging* 1997; 18: S1–2.
- The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. Consensus report of the Working Group on: "Molecular and Biochemical Markers of Alzheimer's Disease". *Neurobiol Aging* 1998; 19: 109–16.
- Trojanowski JQ, Mattson MP. Overview of protein aggregation in single, double, and triple neurodegenerative brain amyloidoses. *Neuromolecular Med* 2003; 4: 1–6.
- Varma AR, Snowden JS, Lloyd JJ, Talbot PR, Mann DM, Neary D. Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 1999; 66: 184–88.

Appendix A. Clinical scenario no. 42

HPI: Sixty-six-year old right-handed woman with 2 years of progressive decline in speech. Initial symptoms included a significant decrease in sentence length and repetitive

performance of routine household chores, as though she had forgotten having done them already. Whereas she was previously an outgoing woman, she became reclusive, shying away from social interactions. After 1 year, she had little spontaneous verbal output. She received speech therapy following a possible transient ischaemic attack characterized by brief unresponsiveness, but derived no benefit. She had a second episode in year 2 consisting of staring followed by unconsciousness. She has stopped cooking and is reluctant to eat unless food placed before her. Tends to wear the same outfit repetitively, and no longer performs housework or uses the phone. When walking to a destination one-fourth mile away, she wandered 3 miles off track.

SH: Prior homemaker and clerk, and married.

FH: Negative for dementia.

Mental status: Quiet with apathetic affect. Did not speak unless specifically questioned. Oriented to city and year but not month. Unable to perform any calculations but recalled 3 out of 3 objects at 5 min after a prolonged registration process. Extremely telegraphic speech. Significant naming difficulties for simple items. Could perform two step and crossed midline commands. Inaccurate when stating age. Formal neuropsychological testing revealed verbal IQ 62, performance IQ 72 and memory quotient 59.

Neuro exam: Possible right facial droop and slight cogwheel rigidity of both upper extremities. Glabellar, snout, grasp and palmomental reflexes all present. Unable to learn tandem gait.

After 2.5 years of symptoms, she is able to say only single words or brief phrases. She has moved to a new home and is able to find her way around the new home. Easily distracted during meals and seems more restless overall. Dresses herself and walks independently, albeit more slowly. Needs assistance with bathing and can help set the table. No crying episodes. On exam, alert and cooperative but does not respond verbally to any direct questions. Able to write out short phrases. Written responses are perseverative. She is able to identify some colours but confabulates with other answers. Unable to point to named body parts or mimic hand movements but can copy a cube. Strong grasp reflexes bilaterally as well as palmomental, glabellar and snout reflexes. Limb reflexes also increased.

After 3 years, she speaks at most one word with prodding. Dresses herself and uses the bathroom appropriately and still assists with some household activities. Walking 1 mile per day. Able to write some notes, but these are frequently meaningless. During the exam, she will inappropriately arise and wander, even when spoken to. Able to close eyes to command but not protrude tongue to command. Does not verbalize during the entire visit. Unable to write her son's name. Slightly stooped gait with decreased associated movements and an extra step when turning. Limb reflexes mildly increased.

After 3.5 years, she is able to recognize familiar persons and cooperate with family members. Appetite good but taking longer to eat. Able to walk within neighbourhood without getting lost. Nocturnal incontinence once per month. Selects own clothing but needs supervision to prevent inappropriate layering. Needs reminders for bathing. Sleeping well. On examination, she is mute and does not respond to simple verbal commands. Able to write her name with difficulty, but perseverates or writes in neologisms to most questions. Unable to copy a figure. Facial expression limited but occasionally smiles. Slightly kyphotic posture. No tremor. Reflexes brisk with palmomental reflexes, increased jaw jerk, snout and sustained glabellar reflexes.

After 4 years, she wanders frequently within the home. Incontinent of urine at night. Able to assist with dressing. Clenches or sits on hands. Has had two additional episodes of 'passing out'. Good appetite but occasionally chokes when putting too much food in her mouth or chewing inadequately. On examination, she remains mute and does not respond to any commands. Fails to make eye contact and is not socially appropriate. She arises spontaneously with a mildly stooped gait and has a lack of facial expression. Grasp and palmomental reflexes prominent bilaterally. Glabellar and jaw jerk also present.

After 4.5 years, she needs occasional assistance with eating but sleeps well. Became lost when walking with her husband outside the home. Frequently incontinent. Rocks back and forth in a chair while seated. Family feels she can still recognize individuals, and she laughs appropriately at jokes. On exam, she remains mute and can follow limited gestured commands. Mild facial masking with cogwheel rigidity and limb hyperreflexia. Dramatic grasp reflexes, prominent snout and glabellar reflexes. Clasps hands in front or behind her when walking and has a moderate kyphosis with slow turns.

After 5 years her husband is dressing and feeding her, but she remains cooperative.

After 6 years, she resides in a nursing home and is walking only with assistance and requires tube feeding. Does not react to other persons. Incontinent. Generally sits with eyes closed.

After 7 years, she requires total care, including for transfers. Opens eyes when name called or shoulder touched but otherwise keeps eyes closed.

She died after 7 years of symptoms.

Symptom checklist

(adapted from Barber *et al.*)

Date _____

Scenario Number _____ or Autopsy Number _____

Site: (circle) UCD U Penn UM

Rater: _____

**Onset of symptom in first third of typical course of progressive dementia
(Circle if symptom present)**

Any memory impairment	–1	Dressing problems	–1
Recent memory loss	–1	Impaired object manipulation	–1
New learning difficulties	–1	Any change in personality	+1
Regularly loses objects	–1	Loss of empathy	+1
Disorientation to place	–1	Inappropriate affect	+1
Paraphasias	–1	Disinhibition	+1
Mutism	+1	Suspiciousness	–1
Any topographic impairment	–1	Distress about handicaps	–1
Navigation difficulties in new environments	–1	Anxiety about handicaps	–1
Navigation difficulties in neighborhood	–1	Loss of confidence	–1
Total this column (range +1 to –9)	_____	Total this column (range +4 to –6)	_____
		Total this section (range +5 to –15)	_____

Symptom Checklist p. 2

Date _____ Scenario Number _____
 or Autopsy Number _____
 Rater: _____

**Onset of symptom in second third of typical course of progressive dementia
(Circle if symptom present, cross section out if scenario doesn't include second third of illness)**

Navigation difficulties within home	–1	Aggression	–1
Impaired facial recognition	–1	Myoclonus	–1
Impaired object recognition	–1		
Impaired object location	–1	Total this section (range 0 to –6)	_____

**Symptoms absent even in last third of the typical course of progressive dementia
(Circle if symptom never observed, cross section out if scenario doesn't include last third of illness)**

Does not regularly lose objects	+1	No disinhibition	–1
No paraphasias	+1	No aggression	+1
No mutism	–1	No suspiciousness	+1
No navigation difficulties in neighborhood	+1	No distress about handicaps	+1
No navigation difficulties in home	+1	No anxiety about handicaps	+1
No impaired object recognition	+1	No loss of confidence	+1
No impaired object location	+1	No myoclonus	+1
No loss of empathy	–1		_____
No inappropriate affect	–1	Total this column (range +3 to –4)	_____
Total this column (range +6 to –3)	_____	Total this section (range +9 to –7)	_____
		Total Symptom Score (range +14 to –28)	_____