Evaluation of a Calibrated $^{18}$F-FDG PET Score as a Biomarker for Progression in Alzheimer Disease and Mild Cognitive Impairment

Karl Herholz$^1$, Sarah Westwood$^1$, Cathleen Haense$^1$, and Graham Dunn$^2$

$^1$Wolfson Molecular Imaging Centre, School of Cancer and Enabling Sciences, University of Manchester, Manchester, United Kingdom; and $^2$School of Community Based Medicine, University of Manchester, Manchester, United Kingdom

Increasingly, clinical trials are being planned in patients with mild cognitive impairment (MCI) to prevent or delay the onset of dementia in Alzheimer disease (AD) by disease-modifying intervention. Inclusion of imaging techniques as biomarkers for patient selection and assessment of outcome is expected to increase trial efficacy. PET using $^{18}$F-FDG provides objective information about the impairment of synaptic function and could, with appropriate standardization, qualify as a biomarker.

Methods: We evaluated a predefined quantitative measure (PET score) that is extracted automatically from $^{18}$F-FDG PET scans using a sample of controls ($n = 44$), patients with MCI ($n = 94$), and patients with mild AD ($n = 40$) from the Alzheimer Disease Neuroimaging Initiative (ADNI). Subjects received 4 scans and clinical assessments over 2 y. Results: PET scores provide much higher test–retest reliability than standard neuropsychologic test scores (Alzheimer’s Disease Assessment Scale–Cognitive [ADAS-cog] and Mini-Mental State Examination) and superior signal strength for measuring progression. At the same time, they are related linearly to ADAS-cog scores, thus providing a valid measure of cognitive impairment. In addition, PET scores at study entry in MCI patients significantly predict clinical progression to dementia with a higher accuracy than Mini-Mental State Examination and ADAS-cog. Conclusion: $^{18}$F-FDG PET scores are a valid imaging biomarker to monitor the progression of MCI to AD. Their superior test–retest reliability and signal strength will allow the reduction in the number of subjects needed or shortening of study duration substantially.

Key Words: neurology; PET; $^{18}$F-FDG; dementia; mild cognitive impairment

DOI: 10.2967/jnumed.111.090902

It is unlikely that drugs or other interventions would be able to reverse the symptoms of manifest dementia in Alzheimer disease (AD) because extensive and irreversible microstructural neuronal changes have already occurred at the time of clinical manifestation of dementia. Thus, there is an urgent need and interest to conduct clinical trials during the prodromal phase in patients with mild cognitive impairment (MCI), with the aim to prevent or at least delay the onset of dementia by disease-modifying intervention (1).

It is currently difficult to perform such trials because the implementation of clinical criteria for MCI and referral pathways that affect patient selection differ widely between centers, resulting in substantial patient sample heterogeneity in multicenter trials and poor comparability of patient samples between trials. Furthermore, most neuropsychologic tests used as standard outcome criteria have substantial measurement variation, and their sensitivity to changes that are typically occurring in MCI patients is limited. Using progression to dementia as the main outcome criterion usually has the disadvantage that it is not completely objective and only a minority of MCI patients will develop clinical dementia within 1–2 y (2). These factors limit the statistical power and thus increase the sample size and associated trial cost. These problems could potentially be overcome by the use of imaging biomarkers as primary outcome parameters with reduced variation. However, for qualification as outcome parameters, their close relationship with clinically relevant outcome has to be demonstrated (3). This should then result in smaller sample size or shorter trial duration without loss of study power.

PET with the widely available tracer $^{18}$F-FDG measures local glucose metabolism as a proxy for neuronal activity at a resting state. Impaired activity in AD is evident as reduced $^{18}$F-FDG uptake predominantly in temporoparietal association areas, including the precuneus and posterior cingulate (4). These changes become detectable in individual subjects as significant deviation from controls 1–2 y before onset of dementia and are closely related to cognitive impairment (5). Compared with MRI morphometry, which is most sensitive in detecting and monitoring hippocampal atrophy and closely related to performance in memory tasks (6), $^{18}$F-FDG PET is more sensitive in detecting neuronal dysfunction in neocortical association areas. The function of these areas is primarily related to cognitive deficits in nonmemory domains such as language and orientation (7),

Received Mar. 24, 2011; revision accepted May 18, 2011.
For correspondence or reprints contact: Karl Herholz, Wolfson Molecular Imaging Centre, University of Manchester, 27 Palatine Rd., Manchester M20 3LJ, U.K.
E-mail: karl.herholz@manchester.ac.uk
COPYRIGHT © 2011 by the Society of Nuclear Medicine, Inc.
which are of particular interest at the stage of transition from a relatively pure memory deficit in MCI to a more extensive cognitive impairment that characterizes dementia. Thus, 18F-FDG PET appears to be particularly well suited for monitoring of progression at that stage, and a preliminary study indicated a substantial increase of study power in clinical trials using regional 18F-FDG uptake as an outcome parameter (8). This potential has recently also been demonstrated by analysis of data from the Alzheimer Disease Neuroimaging Initiative (ADNI), using standard region-based linear modeling (9), and use of a specifically tailored region of interest to maximize sensitivity and specificity (10).

In a previous multicenter study, Herholz et al. (11) developed a measure to quantify the severity of metabolic impairment in AD on 18F-FDG PET scans in an objective manner using an automated procedure, now available commercially as a standard image processing software tool (module PALZ; PMOD Technologies). It has already been applied to cross-sectional ADNI data, demonstrating its robustness and comparability of results in independent large data samples from multiple centers (12). Recently, the large multicenter longitudinal study conducted by the ADNI has compiled a comprehensive sample of healthy controls and patients with MCI and mild AD, providing clinical and 18F-FDG PET data at 4 subsequent times over 2 y. We are therefore now analyzing the properties of the predefined standardized automatic procedure for 18F-FDG PET scans as a candidate biomarker with respect to its reproducibility in controls, correspondence with clinical parameters in patients, signal strength for monitoring progression, and power to predict future progression.

MATERIALS AND METHODS

Data used in the preparation of this article were obtained from the ADNI database (www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations as a $60 million, 5-y public–private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biologic markers, and clinical and neuropsychologic assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness and the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, VA Medical Center and University of California–San Francisco, and ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations. Subjects have been recruited from more than 50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 adults, aged 55–90 y, to participate in the research—approximately 200 cognitively healthy older individuals to be followed for 3 y, 400 people with MCI to be followed for 3 y, and 200 people with early AD to be followed for 2 y. Up-to-date information is provided at www.adni-info.org.

Data for analysis were downloaded from the ADNI Web site in July 2010, including 18F-FDG PET scans and associated clinical and neuropsychologic data. Inclusion criteria for download were completeness of the following items: date of birth, current diagnosis (healthy, MCI, or AD), Mini-Mental State Examination (MMSE), clinical dementia rating, Alzheimer’s Disease Assessment Scale–Cognitive (ADAS-cog), and date of PET scans. Participants were required to have had four 18F-FDG PET scans at baseline, 6 mo, 12 mo, and 24 mo.

PET scans represented the brain activity 30–60 min after injection of 18F-FDG; had been reconstructed using 3-dimensional backprojection, 3-dimensional ordered-subset expectation maximization, or Fourier rebinning/2-dimensional ordered-subset expectation maximization; were scaled to a common global average value; and were reoriented into a standard 160 × 160 × 96 voxel image grid (voxel size, 1.5 × 1.5 × 1.5 mm) along the anterior commissure-posterior commissure (AC-PC) plane and formatted as DICOM or ECAT files. From these scans, we calculated the AD t-sum, as described in previous publications (11,12), using the procedure implemented as module PALZ in the PMOD software package (version 3.2; PMOD Technologies). The AD t-sum indicates the severity of the metabolic decrease in those brain areas that are typically being affected by AD (multimodal association cortices mostly located in the temporal and parietal lobes), including an adjustment for age effects.

In the present study, the AD t-sum was converted into a PET score by reference to its upper normal limit, as determined previously (11), and log transformation to approach a normal distribution of values, according to the following equation:

\[ \text{PETscore} = \log_2 \left( \frac{\text{ADtsum} + 11.089}{11} \right). \]

Results were analyzed using the R software package (R Foundation for Statistical Computing), mainly with regression analysis (function, lm) and a mixed-model ANOVA (function, ezANOVA), as indicated in the text. Random effects and instrumental variable (IV) models were fitted using xtregr and ivregress commands in Stata software (version 11; StataCorp), respectively. Instrumental variable models were fitted by maximum likelihood in Mplus (version 5.21; Muthén & Muthén).

RESULTS

Findings in Diagnostic Groups (at Each Time Point)

The basic demographic data are listed in Table 1. Ages [Table 1] were comparable between groups, and MMSE scores indicate that most AD patients were still only mildly demented at study entry.

As to be expected, the number of patients diagnosed with AD increased over time, whereas there was a decrease of patients who were diagnosed with MCI. The number of controls increased slightly because there were more MCI patients reclassified as controls during follow-up than the reverse. There were no substantial changes of PET scores over time in controls and in subjects diagnosed as MCI at each time point (Table 2)—a result to be expected because subjects with substantial clinical progression moved to the AD group. In contrast, there was a steady increase of PET scores in AD patients, because those who already had AD at entry remained in this group while the disease progressed.

<table>
<thead>
<tr>
<th>PET Score Biomarker for AD Progression</th>
<th>Herholz et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>jnm090902-pm</td>
</tr>
<tr>
<td></td>
<td>7/11/11</td>
</tr>
</tbody>
</table>
The frequency of abnormal PET findings was low (~10%–20%) in healthy controls, 40%–50% in MCI patients, and stably around 85% in AD patients. These proportions were similar at all time points.

When patient groups based on their diagnostic classification at entry without diagnostic reclassification were followed (Table 3), the PET score in AD patients increased by 6.3% at 6 mo, 9.0% at 12 mo, and 25.6% at 24 mo (repeated-measures ANOVA, \( P < 0.0001 \)); the latter 2 times were significant using the Fisher least-significant-difference test. A similar increase, significant already at month 6, was observed in MCI patients by 6.3%, 14.9%, and 28.1% after 6, 12, and 24 mo, respectively, whereas controls remained essentially stable. The progression toward more severe cognitive impairment in MCI and AD patients was also reflected in a decrease of MMSE scores.

**Prediction of Outcome After 24 Months**

Then, subjects were assigned to groups defined by comparison of diagnoses at baseline and at the ultimate follow-up evaluation at 24 mo. Of the 44 controls, 42 remained in that category, and 2 progressed to MCI. Of the 94 patients entering the study as MCI, 7 reverted to control status, 57 remained MCI, and 30 progressed to AD. All patients with manifest AD at baseline remained in that category at follow-up.

PET scores were significantly different between these groups (ANOVA, \( F_{5,173} = 20.479, P < 0.0001 \)); AD patients and subjects progressing to a more severe diagnostic category (controls to MCI and MCI to AD) had significantly higher scores than stable controls and MCI patients (Fig. 1). In contrast, ADAS-cog and MMSE at baseline were different between the main diagnostic groups but did not differ significantly between progressive and nonprogressive MCI.

Subjects who had a PET score at baseline above 1 had a significantly increased risk for progression. The sensitivity to predict progression was 0.57; specificity, 0.67; positive predictive value, 0.45; and negative predictive value, 0.77. The area under the receiver-operating characteristic curve.
Reproducibility and Signal Strength of Measurement

PET scores in stable controls did not change significantly over time in repeated-measures ANOVA (Table 4). On the basis that these subjects are not expected to show any significant change when measured repeatedly, we used them to assess measurement reproducibility. Assuming that the SD for change over 6 mo nearly entirely reflects measurement variation, the figures translate into a coefficient-of-measurement variation of 16%. At months 12 and 24, variation goes up slightly, most likely reflecting the increasing effect of biologic variation as time passes even in controls. The estimate of reliability (intraclass correlation) is 0.922 (95% confidence interval, 0.886–0.958). The signal strength, defined analogous to Cohen d as the difference of mean values relative to measurement variation, for distinction between AD and controls was 10.9 (score difference, 1.11 divided through measurement SD 0.101).

Changes in Progressive Versus Nonprogressive MCI

The increase of PET scores was more closely related to disease progression in MCI patients, as detected clinically by development of dementia during follow-up, than the increase of ADAS-cog scores (Table 6). This difference was most clearly seen at 12 mo, when the increase of PET scores was significantly higher in progressing than in nonprogressing patients, whereas the increase of ADAS-cog was not. In terms of signal strength (Cohen d), the change

**TABLE 3**

<table>
<thead>
<tr>
<th>Diagnosis at entry</th>
<th>Month 6</th>
<th>Month 12</th>
<th>Month 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>-0.068 ± 1.087</td>
<td>0.090 ± 1.326</td>
<td>-0.068 ± 1.265</td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>-0.720 ± 3.224</td>
<td>-1.250 ± 3.384</td>
<td>-1.211 ± 2.950</td>
</tr>
<tr>
<td>PET score</td>
<td>-0.015 ± 0.010</td>
<td>0.022 ± 0.136</td>
<td>0.032 ± 0.166</td>
</tr>
<tr>
<td>MCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>-0.298 ± 2.184</td>
<td>-0.564 ± 2.227</td>
<td>-1.255 ± 3.054</td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>0.791 ± 3.687</td>
<td>1.429 ± 4.254</td>
<td>2.199 ± 4.582</td>
</tr>
<tr>
<td>d (ADAS-cog)</td>
<td>0.215</td>
<td>0.336</td>
<td>0.480</td>
</tr>
<tr>
<td>PET score</td>
<td>0.065 ± 0.235</td>
<td>0.150 ± 0.264</td>
<td>0.286 ± 0.350</td>
</tr>
<tr>
<td>d (PET)</td>
<td>0.276</td>
<td>0.568</td>
<td>0.817</td>
</tr>
<tr>
<td>AD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>-0.750 ± 3.053</td>
<td>-1.975 ± 3.570</td>
<td>-3.350 ± 3.997</td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>1.009 ± 3.528</td>
<td>3.209 ± 5.050</td>
<td>7.835 ± 7.604</td>
</tr>
<tr>
<td>d (ADAS-cog)</td>
<td>0.286</td>
<td>0.635</td>
<td>1.030</td>
</tr>
<tr>
<td>PET score</td>
<td>0.109 ± 0.307</td>
<td>0.175 ± 0.308</td>
<td>0.456 ± 0.417</td>
</tr>
<tr>
<td>d (PET)</td>
<td>0.355</td>
<td>0.568</td>
<td>1.094</td>
</tr>
</tbody>
</table>

Data are mean ± SD and Cohen d as indicated.

**FIGURE 1.** Box plot of baseline PET AD scores (interquartile and full range) for diagnostic groups. AD patients and MCI patients progressing to AD have significantly higher scores than stable subjects (arrows in top insert, P < 0.05 in Tukey multiple comparisons). C = control.
of PET scores as outcome parameter over 12 mo provided nearly the same study power as the ADAS-cog scores over 24 mo.

Relative Calibration of PET Score Against ADAS-cog

There is a close and highly significant correlation between PET scores and ADAS-cog \((r = 0.63, \text{Fig. 3})\) and MMSE scores \((r = -0.63)\) using all available data. Highly significant correlations also exist between the changes of ADAS-cog and PET scores over 24 mo \((r = 0.47)\); they are close, especially in progressing MCI subjects \((r = 0.59)\). Considering baseline data only, to avoid any distortions introduced by practice effects in ADAS-cog, the correlation between PET score and ADAS-cog is 0.55, with the corresponding regression coefficient (effect of PET on ADAS) estimated to be 4.52 using ordinary least squares \((SE, 0.52)\). However, the latter estimate will be attenuated by measurement errors in the PET score. If we allow for measurement error in both variables, then the model is unidentified (in particular, we cannot estimate the required regression coefficient).

Both scores, ADAS-cog and PET, will yield a value of zero in the absence of any abnormality. Although absence of abnormality does not usually occur in actual measure-ments because even healthy controls will usually show some mild abnormalities, it is reasonable to assume a zero intercept in the linear model relating ADAS-cog to PET score. Then, the regression coefficient can be simply estimated by the ratio of the means of the 2 variables \((12.680/1.195 = 10.61)\). Alternatively, IV regression \((13)\) provides an estimate of the regression coefficient without the need for assuming a zero intercept. The resulting IV estimate was 10.58 \((SE, 1.29)\) using the 2 binary dummy variables distinguishing the 3 diagnostic groups as instruments. Repeating the IV regression after replacing the initial diagnosis by baseline MMSE score as the instrument yielded a similar value \((11.93; SE, 1.57)\). Finally, refitting the IV model using maximum likelihood, with an additional zero intercept constraint, yielded a similar estimated value of 10.70 but with a considerably smaller SE \((0.43)\). In our discussion in the “Power Analysis” section, we will therefore assume the true difference in scale is a multiplicative factor of 10.6.

Power Analysis

We now turn to considering recalibrated PET scores by multiplying by a factor of 10.6. Thus, the PET score changes will have approximately the same scale of measurement as those for ADAS-cog. Using the whole sample (regardless of diagnosis), we obtain the changes from baseline to 12 mo and to 24 mo for the 2 measures as shown in Table 7.

As would be expected from the recalibration, the means of changes in PET score and ADAS-cog at each follow-up time are almost the same. A considerable increase in Cohen d for the PET scores, compared with ADAS-cog, is due to less variance in the PET data. At both follow-up times, the variance of the ADAS changes was about 3 times that of the PET score changes. For planning a randomized clinical trial using a simple \(t\) test to compare the outcomes of 2 randomized groups, the sample size \((\text{per group})\) required for a specified significance level and power is proportional to \(\frac{s^2}{d^2}\), where \(s^2\) is the common variance of the outcome assessments and \(d^2\) is square of the true treatment effect to be detected. If (after appropriate calibration) the treatment effect \((\delta)\) is the same for the 2 competing outcome measures, and assuming that the common within-group variances will be similar to those described above (or, at least their ratio will be \(\sim 3\)), then the sample size required for a trial using the ADAS-cog as the primary outcome will

---

TABLE 4

<table>
<thead>
<tr>
<th>Subject no.</th>
<th>Month</th>
<th>(n)</th>
<th>Mean</th>
<th>SD</th>
<th>Change relative to baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>42</td>
<td>0.622</td>
<td>0.363</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>42</td>
<td>0.610</td>
<td>0.384</td>
<td>-0.012</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>42</td>
<td>0.647</td>
<td>0.358</td>
<td>0.026</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>42</td>
<td>0.655</td>
<td>0.376</td>
<td>0.034</td>
</tr>
</tbody>
</table>

TABLE 5

<table>
<thead>
<tr>
<th>Subject no.</th>
<th>Month</th>
<th>(n)</th>
<th>Mean</th>
<th>SD</th>
<th>Change relative to baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>42</td>
<td>7.119</td>
<td>2.856</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>42</td>
<td>6.563</td>
<td>3.117</td>
<td>-0.556</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>42</td>
<td>5.945</td>
<td>3.024</td>
<td>-1.174</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>42</td>
<td>5.739</td>
<td>2.951</td>
<td>-1.38</td>
</tr>
</tbody>
</table>

---

FIGURE 2. Receiver-operating-characteristic curve for prediction of progression of MCI to AD by PET score at entry.
need to be about 3 times that needed if PET scores are used. Alternatively, when using PET scores, one could keep the sample size the same but shorten the follow-up times—the ratio $\frac{\sigma^2}{\delta^2}$ (the square of the inverse of Cohen d) for the PET scores at 12 mo being close to ADAS cog at 24 mo and for PET scores at 6 mo being close to ADAS-cog scores at 12 mo.

**DISCUSSION**

Our analysis is based entirely on a predefined, commercially available, and user-independent procedure for 18F-FDG PET scans. Our analysis is in contrast to that of other studies on imaging biomarkers, in which the image processing procedures under evaluation typically still have been at development stage (10,14,15). The region of interest from which the PET score is being derived had been identified in the original cross-sectional sample by the correlation of those voxels with MMSE (11), thus making its construction well suited for monitoring of disease progression. The robustness of the procedure has also been demonstrated previously in additional cross-sectional samples (12,16). We are therefore confident that the estimate of the variance associated with the PET score measurements described in the present paper is representative for the general application of the technique.

The PET score fulfills several requirements for qualification as a biomarker, as requested by regulators (17). As demonstrated in the present study, PET score provides an objective measure and has a high test–retest reliability, allowing for the assessment of treatment efficacy in a single patient. PET score is representative for the stage of the prodromal and early AD, at which drugs supposed to modify progression should exert their maximum effect. By being closely related to synaptic density and function (18,19), PET score is representative of the supposed mechanism of action of the drug that acts by preserving and maintaining synaptic function while being generic with respect to any molecular mechanisms that might be engaged in this action. As with all biomarkers, the relation to the desired clinical outcome, for example, prevention of cognitive impairment in AD, still needs to be clearly established. Ideally, the evidence would include 2 positive phase 3 trials in which PET scores corresponded with clinical outcomes, but cumulative evidence from correspondence between clinical and PET outcomes in phase 2 trials might be considered as a more realistic pathway (3).

The increase of study power provided by PET scores is primarily based on their lower measurement variability, particularly obvious in stable healthy controls, for whom the coefficient-of-measurement variation for test–retest within 6 mo was only 16%, less than half of that for ADAS-cog, with excellent reliability at an intraclass correlation coefficient of 0.922, compared with 0.472 for ADAS-cog. In an actual clinical trial, the signal change due to clinical disease progression and the associated variation have to be considered, and both tend to increase with trial duration. Thus, a high accuracy of the measured outcome parameter is of particular importance in short-duration trials and at early stages of AD when the biologic signal change is relatively small. Correspondingly, PET scores provided a substantial increase in signal strength in MCI, whereas in AD a clear increase was present only after 6 mo (Cohen d in Table 3).

### TABLE 6

<table>
<thead>
<tr>
<th>Method</th>
<th>Progressive MCI</th>
<th>Nonprogressive MCI</th>
<th>$P$</th>
<th>Cohen d</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 mo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET</td>
<td>0.25 ± 0.31</td>
<td>0.10 ± 0.23</td>
<td>0.0087</td>
<td>0.55</td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>2.3 ± 4.5</td>
<td>1.0 ± 4.1</td>
<td>NS</td>
<td>0.30</td>
</tr>
<tr>
<td>24 mo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET</td>
<td>0.45 ± 0.39</td>
<td>0.21 ± 0.31</td>
<td>0.0014</td>
<td>0.68</td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>4.1 ± 5.1</td>
<td>1.3 ± 4.1</td>
<td>0.0058</td>
<td>0.60</td>
</tr>
</tbody>
</table>

NS = not statistically significant.

Data are mean difference ± SD, compared with baseline.

![FIGURE 3. Scatterplot of corresponding ADAS-cog and PET scores with IV regression line.](1223eteria.png)
At 12 and 24 mo in AD patients, the signal change due to biologic progression and its associated variation probably dominated over measurement-related variation, resulting in similar signal strength for PET scores and ADAS-cog. 

$^{18}$FDG PET has been used in several clinical trials of AD in the past. In most of these trials the emphasis was on examining whether the drug would increase cerebral glucose metabolism as a pharmacodynamic effect of treatment rather than on assessment of progression. Typically these were short trials of a few weeks or up to 3 mo active treatment duration with $^{18}$FDG PET before and during treatment (20–22). Trials that were extended for at least 6 mo also showed evidence of disease progression by further reduction of glucose metabolism in association cortices (23–26). However, these trials had not been designed to assess progression. For demonstration of a disease-modifying effect, $^{18}$FDG PET should be treated like other functional measures. It may therefore require specific trial design, such as slope analysis, randomized start, or randomized withdrawal, which are currently being explored (27–29). There is also the potential for PET to be used as a sensitive and highly reproducible outcome measure in presymptomatic subjects who are at high genetic risk for developing AD (30).

$^{18}$FDG PET as a technical procedure that is independent of language and educational and cultural background may offer an advantage especially in studies performed across multiple countries. Future studies should clarify whether $^{18}$FDG PET could also monitor progression in noncognitive domains, such as behavior. Clinical studies demonstrating metabolic impairment in the prefrontal cortex in the behavioral type of frontotemporal dementia (31) suggest that this is a realistic possibility.

MRI morphometry is another promising imaging biomarker for diagnosis and monitoring of progression (6), and a substantial increase of power, compared with ADAS-cog, has been demonstrated (15,32). In contrast to $^{18}$FDG PET, results are based entirely on structural imaging and therefore are not influenced by actual synaptic function and pharmacodynamic effects, which facilitates the separation of progression from these functional effects. However, local brain volume may be biased by other confounding factors unrelated to disease progression, such as hydration and nutrition status (33). Thus, MRI morphometry and $^{18}$FDG PET should be used as complementary biomarkers to assess structural and functional changes in a comprehensive manner, and that perspective could become particularly comfortable to patients when provided simultaneously in PET/MRI scanners (34).

Amyloid PET offers superior molecular specificity directly related to a major histopathologic marker of AD that has accumulated for many years before actual onset of dementia (35). However, it is currently unclear whether disease progression would be associated with further increase of tracer binding. Initial follow-up studies indicated that there is little further increase after onset of dementia (36), but recent preliminary results from large multicenter studies (ADNI, Australian Imaging, Biomarkers and Lifestyle) indicate further increase. A decrease of Pittsburgh compound B binding has been observed in patients undergoing clinical trials of drugs that remove amyloid from brain (37), but it has not yet been demonstrated that this would be associated with clinical benefit. Amyloid PET also is clearly complementary to functional assessment because its relation to cognitive deficits is rather weak and can usually be demonstrated only when including amyloid-negative subjects who probably do not have AD at any stage (38).

**CONCLUSION**

Our analysis demonstrates the validity of $^{18}$FDG PET scores as an imaging biomarker for clinical trials to prevent dementia in MCI patients. Longitudinal ADNI data indicate that PET scores provide much higher test–retest reliability than ADAS-cog, which is the most frequently used as an outcome measure in dementia trials. By having a close and largely linear relation to ADAS-cog scores, PET scores also provide a valid measure of cognitive impairment. As a measure of disease progression, PET scores may provide a power for 1-y studies in MCI patients similar to what they do.
provide for 2-y studies based on progression of ADAS-cog scores.

DISCLOSURE STATEMENT

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

ACKNOWLEDGMENTS

The study was funded in part by Alzheimer’s Research U.K. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health grant U01 AG024904). ADNI is funded by the National Institute on Aging, by the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc., F. Hoffman-La Roche, and Schering-Plough, Synarc, Inc., as well as nonprofit partners the Alzheimer’s Association and Alzheimer’s Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514, and the Dana Foundation. Data used in preparation of this article were obtained from the ADNI database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Authorship_List.pdf. No other potential conflict of interest relevant to this article was reported.

REFERENCES


