Different definitions of neurodegeneration produce similar amyloid/neurodegeneration biomarker group findings

Clifford R. Jack, Jr.,1 Heather J. Wiste,2 Stephen D. Weigand,2 David S. Knopman,3 Michelle M. Mielke,2 Prashanthi Vemuri,1 Val Lowe,1 Matthew L. Senjem,1 Jeffrey L. Gunter,1 Denise Reyes,1 Mary M. Machulda,4 Rosebud Roberts2 and Ronald C. Petersen3

We recently demonstrated that the frequencies of biomarker groups defined by the presence or absence of both amyloidosis (A+) and neurodegeneration (N+) changed dramatically by age in cognitively non-impaired subjects. Our present objectives were to assess the consequences of defining neurodegeneration in five different ways on the frequency of subjects classified as N+, on the demographic associations with N+, and on amyloidosis and neurodegeneration (A/N) biomarker group frequencies by age. This was a largely cross-sectional observational study of 1331 cognitively non-impaired subjects aged 50–89 drawn from a population-based study of cognitive ageing. We assessed demographic associations with N+, and A/N biomarker group frequencies by age where A+ was defined by amyloid PET and N+ was defined in five different ways: (i) abnormal adjusted hippocampal volume alone; (ii) abnormal Alzheimer’s disease signature cortical thickness alone; (iii) abnormal fluorodeoxyglucose positron emission tomography alone; (iv) abnormal adjusted hippocampal volume or abnormal fluorodeoxyglucose positron emission tomography; and (v) abnormal Alzheimer’s disease signature cortical thickness or abnormal fluorodeoxyglucose positron emission tomography. For each N+ definition, participants were assigned to one of four biomarker groups; A+/N−, A−N+, A−N−, or A+N+. The three continuous individual neurodegeneration measures were moderately correlated (rs = 0.42 to 0.54) but when classified as normal or abnormal had only weak agreement (κ = 0.20 to 0.29). The adjusted hippocampal volume alone definition classified the fewest subjects as N+ while the Alzheimer’s disease signature cortical thickness or abnormal fluorodeoxyglucose positron emission tomography classification classified the most as N+. Across all N+ definitions, N+ subjects tended to be older, more often male and APOE4 carriers, and performed less well on functional status and learning and memory than N− subjects. For all definitions of neurodegeneration, (i) the frequency of A−N− was 100% at age 50 and declined monotonically thereafter; (ii) the frequency of A−N− increased from age 50 to a maximum in the mid-70s and declined thereafter; and3 (iii) the frequency of A−N− (suspected non-Alzheimer’s pathophysiology) and of A+N+ increased monotonically beginning in the mid-50s and mid-60s, respectively. Overall, different neurodegeneration measures provide similar but not completely redundant information. Despite quantitative differences, the overall qualitative pattern of the A−N−, A+N−, A−N+, and A+N+ biomarker group frequency curves by age were similar across the five different definitions of neurodegeneration. We conclude that grouping subjects by amyloidosis and neurodegeneration status (normal/abnormal) is robust to different imaging definitions of neurodegeneration and thus is a useful way for investigators throughout the field to communicate in a common classification framework.
Introduction

The introduction of new diagnostic criteria for Alzheimer’s disease (Albert et al., 2011; Jack et al., 2011a; McKhann et al., 2011; Sperling et al., 2011; Dubois et al., 2014) has focused interest in categorizing individual subjects on the basis of both amyloidosis and neurodegeneration. If individuals are classified as amyloid positive (A+) or negative (A−), and neurodegeneration positive (N+) or negative (N−), then all subjects can be placed into one of four biomarker categories: A−N−, A+N−, A−N+, or A+N+. Relating this classification scheme to the National Institute on Aging–Alzheimer’s Association (NIA-AA) preclinical Alzheimer’s disease staging, A−N− corresponds to stage 0, A+N− to stage 1, and A+N+ to stages 2 plus 3 (Sperling et al., 2011). A−N+ corresponds to suspected non-Alzheimer’s disease pathophysiology (SNAP), a category that falls outside preclinical Alzheimer’s disease and was first described in Jack et al. (2012). In a recent report we observed that in a population-based cohort of cognitively non-impaired subjects aged 50–89 years, the frequencies of the A/N biomarker groups changed dramatically by age (Jack et al., 2014b). In Jack et al. (2014b) we defined A+ by amyloid PET and defined N+ as either abnormal head-size-adjusted hippocampal volume or abnormal fluorodeoxyglucose (FDG) PET, a definition of N+ consistent with the NIA-AA criteria. However, this raised the question whether our findings were unique to the N+ definition we used or were generalizable to different definitions of N+.

The issue of generalizability of neurodegeneration definitions is important. While consensus generally exists that different amyloid PET measures are comparable (Klunk et al., 2015), choices of imaging methods for neurodegeneration may have larger effects (Chetelan et al., 2008; Chetelan, 2013; Wirth et al., 2013b; Alexopoulos et al., 2014; Toledo et al., 2014; Caroli et al., 2015). Moreover, the manner in which different research groups define neurodegeneration varies considerably. Both FDG PET and MRI are widely used measures of neurodegeneration (Jagust et al., 2009) where neurodegeneration is defined as progressive loss of neurons and their processes with a corresponding progressive impairment in neuronal function (Jack and Holtzman, 2013; Jack et al., 2013a). Within MRI, different regions of interest have been used, perhaps the most popular being hippocampal volume and Alzheimer’s disease signature cortical thickness (Davatzikos et al., 2008; Vemuri et al., 2008a; Dickerson et al., 2009; Frisoni et al., 2010; Wirth et al., 2013b). Thus a number of different methods and combinations of methods of defining neurodegeneration by imaging exist. If each of these behaves in a unique manner, then it becomes quite difficult for different investigators to report data and communicate in a common framework. Fragmentation of the field will result, which will hinder progress.

Our present report contains 1331 cognitively non-impaired individuals, 74% of whom appeared in an earlier publication (Jack et al., 2014b). The purpose of our previous report (Jack et al., 2014b) was to describe the frequencies of the four biomarker-defined biomarker groups (defined by amyloid PET and only one of many possible definitions of neurodegeneration) by age. We have two major new objectives in the present study. The first is to assess the effects of different definitions on the number of individuals classified as having abnormal neurodegeneration and on the association between N+ and demographic features, functional status, and memory performance both cross-sectionally and longitudinally. The second was to assess the consequences of defining neurodegeneration in five different ways on A/N biomarker group frequencies by age, among cognitively non-impaired individuals aged 50–89, from a population-based cohort. We demonstrate that while different definitions of neurodegeneration produce some quantitatively different results, qualitative behaviour is largely similar. We conclude that grouping subjects by amyloidosis and neurodegeneration status (normal/abnormal) is robust to different imaging definitions of neurodegeneration and thus is a useful way for investigators throughout the field to communicate in a common classification framework.

Materials and methods

Subject recruitment and inclusion criteria

All 1331 subjects in this study were participants in the Mayo Clinic Study of Aging (MCSA). The MCSA is a population-based study of cognitive ageing among Olmsted County, MN, USA residents (Roberts et al., 2008). The Olmsted County population was enumerated for 50–89 year olds. From this enumeration, subjects were selected for recruitment using a
5-year age- and sex-stratified random sampling strategy such that males and females were equally represented in each 5-year age category. Subjects selected for recruitment were invited to participate in the MCSA and those without a medical contraindication were invited to participate in imaging studies. Of the subjects that participated in the MCSA clinical visit, 71% consented to the complete imaging protocol: MRI, FDG PET and amyloid PET. To be included in the present study subjects must have been judged clinically to have no cognitive impairment for age (i.e. neither mild cognitive impairment nor dementia) based on psychometric testing and independent evaluations by a study coordinator, neuropsychologist and a physician (Roberts et al., 2008). We therefore label subjects in this study as cognitively ‘non-impaired’. Subjects also had to have undergone amyloid PET, FDG PET, and MRI. A total of 1331 participants met these inclusion criteria. The imaging studies were obtained from March 2006 to April 2015. Clinical status was assessed near the time of imaging for all subjects. The MRI was performed at a median of 2 months after the clinical visit (range: 0 to 5 months) and the PET imaging was performed at a median of 3 months (range: 0 to 6 months) after the clinical visit. Ninety-eight per cent of subjects in this study were Caucasian, 0.5% Asian, 0.2% Black/African American, 0.7% identify with more than one race, and 0.4% were unknown/not disclosed. Ethnicity was non-Hispanic in 99%, Hispanic in 0.2%, and unknown/not disclosed in 0.5% of subjects.

We assessed the effect of different definitions of neurodegeneration on the associations with functional status and memory performance both cross-sectionally and longitudinally. Functional status was assessed by the Mini-Mental State Examination (MMSE) and memory and learning performance was assessed by the Auditory Verbal Learning Test (AVLT). Longitudinal change in learning and memory performance from baseline was assessed by the AVLT. We used the sum of trials 1–5 plus immediate and delayed recall (total possible score = 105) as the AVLT metric.

**Standard protocol approvals, registrations and patient consents**

These studies were approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards. Written informed consent was obtained from all participants.

**Imaging methods**

Amyloid PET imaging was performed with Pittsburgh Compound B (Klunk et al., 2004) and FDG PET was obtained on the same day. CT was obtained for attenuation correction. Amyloid PET images were acquired from 40–60 min and FDG from 30–40 min after injection. Amyloid PET and FDG PET were analysed with our in-house fully automated image processing pipeline (Senjem et al., 2005) where image voxel values are extracted from automatically labelled regions of interest propagated from an MRI template. An amyloid PET standardized uptake value ratio (SUVR) was formed from the prefrontal, orbitofrontal, parietal, temporal, anterior cingulate, and posterior cingulate/precuneus regions of interest normalized to the cerebellum (grey plus white matter). The data were partial volume corrected for voxel CSF content using segmented co-registered MRI. An Alzheimer’s disease-characteristic FDG PET SUVR was formed from the angular gyrus, posterior cingulate, and inferior temporal cortical regions of interest normalized to pons and vermis (Landau et al., 2011). FDG PET data were not partial volume corrected. We and others have reported previously that diagnostic performance is slightly better if amyloid PET is partial volume corrected (Lowe et al., 2009; Su et al., 2015), and is much better if FDG PET is not partial volume corrected (Lowe et al., 2009; Curiati et al., 2011). Consequently these are the methods we used in the present analysis.

MRI was performed on one of three 3 T systems from the same vendor. Two MRI measures were used. We summed right and left hippocampal volumes from Freesurfer (v 5.3), and adjusted them for total intracranial volume (TIV) by calculating the residual from a linear regression of hippocampal volume (y) versus intracranial volume (x) among 133 cognitively normal subjects aged 30 to 59 (described in Jack et al., 2014a). Adjusted hippocampal volume can be interpreted as the deviation in cm$^3$ in a subject’s hippocampal volume from what is expected given their TIV. The second MRI measure was an Alzheimer’s disease signature cortical thickness measure composed of the following individual cortical thickness regions of interest: entorhinal, inferior temporal, middle temporal, and fusiform. In-house evaluation indicated that the Alzheimer’s disease signature composite cortical thickness measure was not correlated with TIV (Spearman rank correlation $r_s = −0.09$, $P = 0.16$) in cognitively non-impaired individuals aged 50–60 (a subgroup chosen in which we are reasonably certain subjects do not harbour latent age-related pathology), whereas hippocampal volume and TIV were strongly correlated ($r_s = 0.62$, $P < 0.001$). We therefore used TIV-adjusted hippocampal volume, but did not adjust the Alzheimer’s disease signature cortical thickness measure for TIV an approach adopted by other groups (Dickerson et al., 2009).

**Definition of abnormality**

The normal/abnormal cut point for the PET and MRI measures was set at the 90th percentile (mild end of the range) of a group of 75 subjects with Alzheimer’s disease dementia from the Mayo Clinic using the same approach as Jack et al. (2012) (Supplementary Fig. 1). Abnormal adjusted hippocampal volume was defined as $<−2.40 \text{ cm}^3$; abnormal FDG PET SUVR as $<1.32$; and abnormal Alzheimer’s disease signature cortical thickness as $<2.74 \text{ mm}$. Abnormal amyloid PET was defined as SUVR $>1.40$, which was recently validated by autopsy correlation with Thal amyloid phase (Murray et al., 2015).

**Definition of neurodegeneration and A/N biomarker groups**

$N^+$ was defined in five different ways: (i) abnormal adjusted hippocampal volume alone; (ii) abnormal Alzheimer’s disease signature cortical thickness alone; (iii) abnormal FDG PET alone; (iv) abnormal adjusted hippocampal volume or abnormal FDG (HV/FDG) (as in Jack et al., 2012, 2014b); and (v) abnormal Alzheimer’s disease signature cortical thickness or abnormal FDG PET abbreviated as ADsig/FDG.
Other combinations are possible but we limited the evaluation to these five to keep the data presentation manageable.

For each N+ definition, participants were assigned to an A/N biomarker group with A−N− denoting normal amyloid PET and normal neurodegeneration, A+N− denoting abnormal amyloid PET but no neurodegeneration, A−N+ denoting normal amyloid PET but abnormal neurodegeneration, and A+N+ denoting abnormal amyloid PET and abnormal neurodegeneration.

### Statistical analysis

We summarize associations between continuous MRI or FDG measures using Spearman’s rank correlation (\(r_s\)) and summarize agreement in terms of abnormality using Cohen’s kappa statistic (\(\kappa\)). Differences in the overall proportion of subjects classified as N+ by the five neurodegeneration definitions were tested using logistic regression with generalized estimating equations (GEE) to compare N+ proportions across methods while accounting for the five repeated measures for each subject. Similarly, differences in demographic characteristics such as average age or proportion male by neurodegeneration status and definition were tested using linear or logistic regression with GEEs. For these models, we included neurodegeneration status (N− or N+), neurodegeneration definition (adjusted hippocampal volume alone, Alzheimer’s disease signature thickness alone, etc.), and the interaction between the two. All models were adjusted for age, sex, and education when applicable. We report an overall \(P\)-value for differences by N+ versus N− across all neurodegeneration definitions and a \(P\)-value indicating whether the difference between N− and N+ varies by definition (i.e. a test of interaction).

In a subset of 955 subjects who had at least one follow-up AVLT evaluation after imaging, we fit a separate linear mixed effects regression model with subject-specific intercepts and slopes for each N+ definition. Each model included age \(\times\) time and N+ \(\times\) time interactions, which allowed baseline values and rates of decline to differ by age and whether the subject was N+ or N−.

To summarize the frequencies of N+ by age, we used an approach described previously (Jack et al., 2014b). Five separate logistic regression models (one for each N+ definition) were used to model the log odds of N+ by age. The estimates were then back-transformed using the inverse logit transformation, \(\exp(x)/(1 + \exp(x))\), and multiplied by 100 to obtain the estimated frequency or percentage of subjects deemed as N+ by age. Age was modelled using a restricted cubic spline having knots at ages 60, 70, and 80 years to allow for a more flexible relationship between N+ and age. We compared differences in N+ frequencies across definitions at each age using bootstrap resampling. For each of the 5000 bootstrap samples, the five separate logistic regression models were refit and pairwise differences in frequencies between each method were estimated for each integer age between 50 and 89. Ninety-five per cent confidence intervals for the N+ definition differences were formed by calculating the 2.5th and the 97.5th quantiles of these differences. We plot these confidence intervals and define significance based on whether the 95% confidence interval excludes the null value, \(y = 0\). For each N+ definition, we model A/N biomarker frequencies as previously reported (Jack et al., 2014b), using multinomial models to estimate the group wise frequencies as a function of age.

### Results

Demographic information for all subjects is displayed in Table 1.

#### Grouping subjects using five definitions of neurodegeneration only: associations of neurodegeneration with demographics

Statue plots comparing the two continuous MRI measures and the continuous FDG measure among all subjects reveal the following pairwise correlations (Fig. 1): Alzheimer’s disease signature thickness versus adjusted hippocampal volume (\(r_s = 0.54\)), adjusted hippocampal volume versus FDG (\(r_s = 0.47\)), Alzheimer’s disease signature thickness versus FDG PET (\(r_s = 0.42\)), all \(P < 0.001\). When each value was scored as N+ or N−; however, there was only weak agreement among these measures with adjusted hippocampal volume and Alzheimer’s disease signature thickness showing the highest agreement (\(\kappa = 0.29\)), followed by Alzheimer’s disease signature thickness and FDG (\(\kappa = 0.25\)), then adjusted hippocampal volume and FDG (\(\kappa = 0.20\)). Figure 2 shows the overlap in classifying subjects as N+ by the three measures, whereas 69% of subjects were normal by all three measures; only 3% of subjects were abnormal by all three measures.

When averaged over all ages, neurodegeneration was most common when defined as ADsig/FDG and least common when defined as adjusted hippocampal volume alone (Fig. 3). N+ subjects tended to be older (\(P < 0.001\)), more often male (\(P < 0.001\)) and APOE4 carriers (\(P = 0.003\)), nominally less educated (\(P = 0.03\)), and performed less well on the MMSE than N− subjects (Fig. 3, \(P < 0.001\)). These observations were generally similar across all N+ definitions. However, the adjusted hippocampal volume alone N+ group had a greater proportion of males, were slightly older and performed slightly worse on MMSE than N− subjects with other definitions. N+ subjects performed worse on the AVLT at baseline using all five definitions of N+ (\(P < 0.001\) to \(P = 0.002\)), although the baseline differences were greatest using the adjusted hippocampal volume alone definition. In longitudinal analyses among the subset of 955 subjects, memory declined more in those who were N+ compared to N− for all definitions (\(P < 0.001\) to \(P = 0.009\)) except the FDG alone definition which showed a similar but non-significant trend (\(P = 0.09\)). The difference in decline between N+ and N− was comparatively greater using the adjusted hippocampal volume alone and Alzheimer’s disease signature thickness alone measures (Fig. 4).

The frequency of N+ increased from 0 at age 50 monotonically among males and females separately and combined for all five definitions of neurodegeneration (Fig. 5). The greatest sex effect observed was with the adjusted hippocampal volume alone and HVa/FDG definitions of N+.
where for both (particularly adjusted hippocampal volume alone) the increase in frequency of N+ with age was more pronounced among males than females (Fig. 5).

Pairwise difference plots by age among the five different definitions of neurodegeneration illustrate significant effects of the definition of neurodegeneration (Supplementary Fig. 2). From age 60 and above the frequency of N+ was highest when defined by ADsigFDG. Defining N+ as adjusted hippocampal volume alone resulted in the lowest frequencies of N+. The other definitions occupied a middle ground, although from around age 70 onward frequencies of N+ were higher when defined by HVaFDG compared to abnormal Alzheimer’s disease signature cortical thickness alone or FDG alone.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1331</td>
</tr>
<tr>
<td>Age, years, median (IQR)</td>
<td>71 (63, 77)</td>
</tr>
<tr>
<td>Male gender, n (%)</td>
<td>699 (53%)</td>
</tr>
<tr>
<td>Education, years, median (IQR)</td>
<td>15 (12, 16)</td>
</tr>
<tr>
<td>APOE e+ positive, n (%)</td>
<td>353 (27%)</td>
</tr>
<tr>
<td>MMSE, median (IQR)</td>
<td>29 (28, 29)</td>
</tr>
<tr>
<td>AVLT sum of trials, median (IQR)</td>
<td>61 (50, 73)</td>
</tr>
<tr>
<td>Amyloid PET, SUVR</td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.33 (1.27, 1.43)</td>
</tr>
<tr>
<td>Abnormal, n (%)</td>
<td>403 (30%)</td>
</tr>
<tr>
<td>FDG PET, SUVR</td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.47 (1.37, 1.58)</td>
</tr>
<tr>
<td>Abnormal, n (%)</td>
<td>237 (18%)</td>
</tr>
<tr>
<td>HVa, cm³ median (IQR)</td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>−.085 (−.1.56, −.20)</td>
</tr>
<tr>
<td>Abnormal, n (%)</td>
<td>102 (8%)</td>
</tr>
<tr>
<td>Alzheimer’s disease signature</td>
<td></td>
</tr>
<tr>
<td>thickness, mm, median (IQR)</td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2.89 (2.79, 2.99)</td>
</tr>
<tr>
<td>Abnormal, n (%)</td>
<td>233 (18%)</td>
</tr>
</tbody>
</table>

IQR = interquartile range; HVa = TIV adjusted hippocampal volume; SUVR = standardize uptake value ratio.

AVLT is the sum of trials 1–5 plus immediate and delayed recall (total possible score 105).

Grouping by amyloidosis and neurodegeneration: A/N group frequencies for different neurodegeneration definitions

The estimated frequency of subjects in each A/N biomarker group by age broken out by the five different definitions of neurodegeneration is shown in Fig. 6. Figure 7 shows the same estimates but is partitioned to emphasize comparison of the age-trend plots across the five neurodegeneration definitions within each A/N biomarker group. The age trends within each A/N biomarker group are similar across the different definitions of N+. For all five definitions of N+, the frequency of A−N− was 100% at age 50 and declined monotonically thereafter. The frequency of A+N− increased from ~0% at age 50 to a maximum in the early to mid-70s and then declined thereafter. However, for the adjusted hippocampal volume alone definition of N+, the data are also consistent with a plateau in the frequency of A+N− at older ages. The frequencies of A−N+ (SNAP) and A+N+ were both 0% at age 50 and increased monotonically from the mid-50s and mid-60s, respectively, although for the ADsigFDG definition, the data are also consistent with a plateau in the frequency of A−N+ at the oldest ages.

Despite this general pattern, some quantitative differences in A/N frequency arise depending on the definition of N+ leading to differences in the relative A/N frequencies at a given age among the different neurodegeneration definitions (Figs 6 and 7). For example, using the ADsigFDG definition, A+N+ was most frequent at age 89 and A−N− and A+N− least frequent. However, when using the adjusted hippocampal volume alone definition, A−N− and A+N− were the most frequent biomarker groups at age 89 and A−N+ and A+N+ the least frequent. A/N frequencies were generally similar using the other three definitions of N+ (Alzheimer’s disease signature thickness alone, FDG alone, and HVaFDG).

Discussion

Our primary findings were: (i) the three continuous individual neurodegeneration measures were moderately correlated but when classified as normal or abnormal had only weak agreement. The adjusted hippocampal volume alone definition classified the fewest subjects as N+ while the ADsigFDG definition classified the most subjects as N+. (ii) Across all N+ definitions, N+ subjects tended to be older, more often male and APOE4 carriers, nominally less educated, and performed less well on the MMSE and AVLT at baseline and had greater decline in AVLT over time than N− subjects. (iii) Despite quantitative differences the overall qualitative pattern of the A−N−, A+N−, A−N+, and A+N+ frequency curves by age were similar across the five different definitions of N+.

The issue of whether different measures of neurodegeneration provide redundant or independent information has been raised by several research groups recently (Chetelat et al., 2008; Chetelat, 2013; Wirth et al., 2013b; Alexopoulos et al., 2014; Toledo et al., 2014; Caroli et al., 2015). The first major objective of this study was therefore to assess the consequences of defining N+ in different ways on the frequency of individuals classified as N+ and the associations with neurodegeneration status and demographic characteristics. The three individual neurodegeneration measures had only weak agreement when subjects were classified as either N+ or N−. When considered as pairs, Alzheimer’s disease signature thickness and adjusted hippocampal volume had the most overlapping...
(redundant) information whereas Alzheimer’s disease signature thickness and FDG PET provided the most independent information. Among the three individual neurodegenerative measures, Alzheimer’s disease signature thickness and FDG alone classified the largest number of subjects as N+ and adjusted hippocampal volume the fewest. Not surprisingly, neurodegeneration was most frequent when defined by the combined ADsig/FDG measure. Thus these neurodegeneration measures provide similar but not completely redundant information.

It may seem counterintuitive that different individual neurodegeneration measures classify different numbers of subjects as N+ given that the cut-points for all three individual neurodegeneration measures were defined in an identical manner among the same group of Alzheimer’s disease subjects (Jack et al., 2012, 2014b) (Supplementary Fig. 1). The two most logical explanations are that (i) the three measures are sensitive to different aspects of pathology and these pathologies are not all concordant (Chetelat et al., 2008; Caroli et al., 2015); or (ii) that the severity of the three neurodegenerative measures in the group of subjects with Alzheimer’s disease dementia in which the cut-off points were defined does not translate directly when applied to cognitively non-impaired individuals. For example, some neurodegenerative measures might change earlier in the disease, or differences in detection sensitivity may exist such that measures can be differentially affected in various stages of disease progression (Chetelat et al., 2010; Villain et al., 2010; Jack et al., 2011b; Benzinger et al., 2013; Villemagne et al., 2013). Associations of N+ with demographic variables were similar in most respects across all N+ definitions. N+ subjects tended to be older, more often male and APOE4 carriers, less educated, and performed less well on the MMSE and AVLT at baseline than N− subjects (Fig. 3). However, the adjusted hippocampal volume alone definition of N+ stood apart from the others in several ways. N+ subjects defined by adjusted hippocampal volume alone were slightly older, performed worse on the MMSE and AVLT at baseline, declined more over time on the AVLT, and had a markedly higher proportion of males than N+ subjects by other definitions (Figs 3 and 4). Also, the increasing frequency of N+ with age was more pronounced among males than females with the adjusted hippocampal volume definition of N+ (Fig. 5). This sex-related behaviour of the adjusted hippocampal volume measure could (i) reflect biological underpinnings (DeCarli et al., 2005); (ii) be a feature of the method used for adjusting raw hippocampal volume for head size; or (iii) be a combination of both.
Hippocampal volume is highly correlated with head size. More generally, cortical grey matter volume and surface area scale with head size whereas cortical thickness does not (Barnes et al., 2010; Winkler et al., 2010). We and others have traditionally used regression methods to adjust raw hippocampal volume for intersubject differences in head size (Jack et al., 1989, 2014b, 2015; Barnes et al., 2010; Mormino et al., 2014a). However, males tend to have larger brains than females (Jack et al., 1995; Barnes et al., 2010), and it may be that regression methods can not...
completely correct for the effect of head size on hippocampal volume in precisely the same manner for males and females over the entire adult age spectrum.

Subjects with neurodegeneration perform worse on the AVLT at baseline and decline more over time than those without (Fig. 4), which validates the meaningfulness of measures of neurodegeneration. N+ subjects defined by adjusted hippocampal volume alone and Alzheimer’s disease signature thickness alone declined somewhat more on the AVLT longitudinally than N+ subjects using other definitions (Fig. 4). In the case of the adjusted hippocampal volume, this association could be ascribed to the adjusted hippocampal volume classifying fewer people as N+ but who have more advanced neurodegenerative disease at baseline and thus more likely to decline cognitively over time compared to other neurodegenerative measures. However, the sensitivity of the Alzheimer’s disease signature thickness measure to AVLT change is not likely due to

Figure 4 Estimated mean baseline AVLT sum of trials and annual change in AVLT sum of trials by neurodegeneration status for the five definitions. Estimates and 95% confidence intervals are from a linear mixed effects model and indicate the mean for an 80-year-old with 15 years of education averaging over males and females. All differences between N− and N+ are significant at the 0.01 level except for rate differences in the FDG alone model (P = 0.09). AVLT sum of trials is the sum of trials 1–5, immediate, and delayed recall (a total score of 105 words is possible). HVa = adjusted hippocampal volume; ADsig = Alzheimer’s disease signature thickness.

Figure 5 Estimated frequency of abnormal neurodegeneration by age for each of the five neurodegeneration definitions. Estimates on the frequency or percentage scale are from an inverse-logit transformation of the log odds obtained from separate logistic regression models. We allow for non-linearity in age by fitting it as a restricted cubic spline with knots at 60, 70, and 80 years. Estimates are shown separately for females (A) and males (B) and also after adjusting for sex (C). C, therefore, represents an average frequency over males and females. Abnormal adjusted hippocampal volume or abnormal FDG is abbreviated as HVa/FDG; and abnormal Alzheimer’s disease signature cortical thickness or abnormal FDG PET is abbreviated as ADsig/FDG.
selective capture of the most impaired people since Alzheimer’s disease signature thickness alone and FDG alone had similar N+ frequencies and FDG was less associated with decline on the AVLT. Thus Alzheimer’s disease signature thickness may simply be more sensitive to impending memory decline than FDG.

The fact that the different neurodegenerative measures behaved differently in some ways impacts choices...
investigators might make when selecting measures for specific purposes. In situations where it is desirable to use a neurodegeneration measure that captures the greatest numbers of individuals the ADsig/FDG metric might be preferred. When classifying individuals where mechanistic specificity is at a premium, one might consider the three individual measures independently (Chetelat et al., 2008). The adjusted hippocampal volume has somewhat different qualities than the other measures, it correlates well with learning and memory cross-sectionally and longitudinally, but also has a stronger relationship with male sex than other measures. Thus Alzheimer’s disease signature thickness may be preferable to adjusted hippocampal volume for many uses as it does not have such a strong relationship with male sex, does not require adjustment for head size, captures as many people as FDG, and correlates well with change over time on the AVLT.

We use the term ‘neurodegeneration’ to describe hippocampal and cortical atrophy and hypometabolism. The pathological substrates of atrophy on MRI have been more thoroughly studied than for FDG PET. Imaging autopsy correlation studies show that atrophy on MRI occurs in a wide variety of pathological conditions (Jack et al., 2002; Barkhof et al., 2007; Di Paola et al., 2008; Vemuri et al., 2008b; Whitwell et al., 2008, 2010; Josephs et al., 2014), which include Alzheimer’s disease but also cerebrovascular disease, hippocampal sclerosis, frontotemporal lobar degeneration syndromes, traumatic encephalopathy, agyrophilic grain disease and perhaps ageing itself without specific pathology (Fjell et al., 2013; Jagust, 2013). Hypometabolism occurs in many of these same disorders (Josephs et al., 2010; Tsitsopoulos and Marklund, 2013; Wirth et al., 2013b). Our subjects were cognitively non-impaired and many of these pathological conditions are associated with cognitive impairment, however, each must exist in a preclinical stage in individuals who later become impaired with the passage of time. If one defines neurodegeneration as progressive loss of neurons and their processes with a corresponding progressive impairment in neuronal function (Jack and Holtzman, 2013; Jack et al.,

Figure 7  Estimated frequency of subjects broken out into each A/N biomarker group by age for the five different definitions of neurodegeneration. Estimates are from a multinomial model adjusted for sex and therefore, the curves represent an average frequency over males and females. Non-linearity in age was allowed in the model by fitting age as a spline with knots at 60, 70, and 80 years. Abnormal adjusted hippocampal volume or abnormal FDG is abbreviated as HVa/FDG; and abnormal Alzheimer’s disease signature cortical thickness or abnormal FDG PET is abbreviated as ADsig/FDG.
2013a), then we believe that the atrophy and hypometabolism are reasonably linked with most age-related substrates of neurodegeneration.

One might ask in addition to neurodegeneration, why not also evaluate the effects of different definitions of amyloid positivity. The two measures are quite different. Amyloid biomarkers are highly specific indicators of the Alzheimer’s disease process whereas neurodegenerative markers are not disease-specific but are sensitive indicators of severity/stage (Jack and Holtzman, 2013; Jack et al., 2013a). We did not have the available data to assess CSF amyloid-β42 in our sample of subjects. However, a recent large meta-analysis successfully combined amyloid PET and CSF amyloid-β42 data, where individual values were scored as normal/abnormal, successfully across many sites without significant site (and hence methodological) effects (Jansen et al., 2015).

The second major objective of this study was to assess the consequences of defining N+ in different ways on amyloidosis and neurodegeneration (A/N) biomarker group frequencies by age. We previously (Jack et al., 2014b) described A/N biomarker frequency trends by age when defining N + as abnormal HVa/FDG, a definition consistent with the NIA-AA criteria for preclinical Alzheimer’s disease (Sperling et al., 2011). We found the frequency of A−N− declined monotonically after age 50. The frequency of A+N− increased from age 50 to a maximum in the mid-70s and declined thereafter. The frequency of A−N+ (SNAP) and A+N+ increased monotonically. The current study confirms those general age-related patterns with several alternative definitions of N+ , which supports the validity of our original conclusions (Figs 6 and 7). The notable changes in cross-sectional A/N group frequencies with age are most logically explained by movement of individual subjects from one biomarker group to another with advancing age. Successful ageing, defined as remaining free of clinically significant cognitive impairment, therefore appears to be characterized by progressive movement of subjects from less to more advanced biomarker abnormality groups. These imaging abnormalities seem to be almost an inevitable consequence of ageing with only ~13% of subjects remaining in the A−N− group at age 89 using the ADsig/FDG definition of N+ .

The frequency of A+N− increased until the mid-70s and then declined with all definitions of N+. One possible explanation offered earlier (Jack et al., 2014b) is that the decrease in frequency of A+N− at ages beyond the mid-70s is consistent with a specific mechanistic sequence in which amyloidosis induces neurodegeneration, an idea strongly supported by the genetics of Alzheimer’s disease (Goate et al., 1991; Levy-Lahad et al., 1995; Sherrington et al., 1995; Jonsson et al., 2012) and biomarker studies in autosomal dominant Alzheimer’s disease (Bateman et al., 2012; Benzinger et al., 2013; Fagan et al., 2014; Fleisher et al., 2015). Thus those who become A+N− in their 60s or early 70s tend to become A+N+ in their late 70s and beyond rather than remain A+N− indefinitely. This corresponds to movement of subjects from preclinical Alzheimer’s disease stage 1 to stage 2 or 3 (Sperling et al., 2011). In contrast, the A−N+ (SNAP) (Jack et al., 2012) frequency continues to increase, which implies that in some individuals neurodegeneration in the absence of amyloidosis may be an end state with no biological propulsion out of this biomarker state. These contrasting population frequency patterns by age for A−N+ versus A+N−, regardless how neurodegeneration was defined, strengthens our contention that A+N− and A−N+ represent different underlying pathological entities (Jack et al., 2014b). We proposed that A+N− represents preclinical Alzheimer’s disease while A−N+ (SNAP) represents a pathologically heterogeneous collection of non-Alzheimer’s disease conditions (Jack et al., 2012; Knopman et al., 2012, 2013). Others have also found amyloid-β independent Alzheimer’s disease-like neurodegenerative biomarker abnormalities in cognitively normal (Jack et al., 2013b; van Harten et al., 2013; Vos et al., 2013; Wirth et al., 2013a, 2013b; Mormino et al., 2014b; Toledo et al., 2014) and mildly impaired subjects (Duara et al., 2013; Petersen et al., 2013; Prestia et al., 2013; Caroli et al., 2015) consistent with this concept of SNAP (Jack et al., 2012).

**Funding**

This study was sponsored by the National Institute on Aging (R01-AG011378, RO1 AG041851, U01 AG06786); the Alexander Family Professorship of Alzheimer’s Disease Research. Dr Jack has provided consulting services for Eli Lilly. He receives research funding from the NIH (R01-AG011378, U01-AG024904, RO1 AG041851, R01 AG37551, R01AG043392, U01-AG06786), and the Alexander Family Alzheimer’s Disease Research Professorship of the Mayo Clinic. Dr. Knopman serves as Deputy Editor for Neurology®; serves on a Data Safety Monitoring Board for Lundbeck Pharmaceuticals and for the DIAN study; is an investigator in clinical trials sponsored by TauRX Pharmaceuticals, Lilly Pharmaceuticals and the Alzheimer’s Disease Cooperative Study; and receives research support from the NIH. Dr Lowe is a consultant for Bayer Schering Pharma and receives research support from GE Healthcare, Siemens Molecular Imaging, AVID Radiopharmaceuticals, the NIH (NIA, NCI), the Elsie and Marvin Dekelboum Family Foundation, the MN Partnership for Biotechnology and Medical Genomics, and the Leukemia & Lymphoma Society. Dr Machulda receives research support from the NIH (U01 AG06786; P50 AG16574; R01 DC 12519). Dr Petersen serves on a Data Monitoring Committee for Pfizer Inc. and Janssen Alzheimer Immunotherap and is a consultant for Roche, Inc., Merck, Inc., and Genentech, Inc.; receives royalties from publishing Mild Cognitive Impairment (Oxford University Press, 2003); and receives research support from the NIH [P50-AG016574 (PI) and...
U01-AG006786 (PI), R01-AG011378 (Co-I), and U01-024904 (Co-I). The other authors report no disclosures.

Supplementary material
Supplementary material is available at Brain online.

References


