ELSEVIER

Alzheimer's & Dementia ■ (2016) 1-12

Alzheimer's & Dementia

Featured Article

Defining imaging biomarker cut points for brain aging and Alzheimer's disease

Clifford R. Jack, Jr.,[a,*], Heather J. Wiste[b], Stephen D. Weigand[b], Terry M. Therneau[b], Val J. Lowe[c], David S. Knopman[d], Jeffrey L. Gunter[e], Matthew L. Senjem[e], David T. Jones[d], Kejal Kantarci[a], Mary M. Machulda[f], Michelle M. Mielke[b], Rosebud O. Roberts[b], Prashanthi Vemuri[a], Denise A. Reyes[a], Ronald C. Petersen[d]

[a]*Department of Radiology, Mayo Clinic, Rochester, MN, USA*
[b]*Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA*
[c]*Department of Nuclear Medicine, Mayo Clinic, Rochester, MN, USA*
[d]*Department of Neurology, Mayo Clinic, Rochester, MN, USA*
[e]*Department of Information Technology, Mayo Clinic, Rochester, MN, USA*
[f]*Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA*

Abstract

Introduction: Our goal was to develop cut points for amyloid positron emission tomography (PET), tau PET, flouro-deoxyglucose (FDG) PET, and MRI cortical thickness.
Methods: We examined five methods for determining cut points.
Results: The reliable worsening method produced a cut point only for amyloid PET. The specificity, sensitivity, and accuracy of cognitively impaired versus young clinically normal (CN) methods labeled the most people abnormal and all gave similar cut points for tau PET, FDG PET, and cortical thickness. Cut points defined using the accuracy of cognitively impaired versus age-matched CN method labeled fewer people abnormal.
Discussion: In the future, we will use a single cut point for amyloid PET (standardized uptake value ratio, 1.42; centiloid, 19) based on the reliable worsening cut point method. We will base lenient cut points for tau PET, FDG PET, and cortical thickness on the accuracy of cognitively impaired versus young CN method and base conservative cut points on the accuracy of cognitively impaired versus age-matched CN method.

Keywords: Alzheimer's disease; Alzheimer's imaging; Alzheimer's MRI; Amyloid PET; Tau PET; FDG PET; Alzheimer's biomarkers; Quantitative imaging

1. Introduction

Imaging and biofluid biomarkers of Alzheimer's disease (AD) are increasingly important to the study of brain aging and dementia. Although every biomarker exists on a continuum, dichotomizing biomarker values is necessary in certain situations. Clinical trials require a normal/abnormal classification when a biomarker is used to determine eligibility [1,2]. Additionally, modern criteria for AD across the cognitive spectrum label an individual's biomarker as normal or abnormal [3–7]. The goal of our study was to develop amyloid positron emission tomography (PET), tau PET, flouro-deoxyglucose (FDG) PET, and structural magnetic resonance imaging (MRI) biomarker cut points.

In brain aging and dementia research, defining a normal/abnormal cut point for quantitative amyloid PET has received

*Corresponding author. Tel.: 507-284-7096; Fax: 507-284-9778.
E-mail address: jack.clifford@mayo.edu

http://dx.doi.org/10.1016/j.jalz.2016.08.005

significant attention. Various methods have been used [8–15] including the 10th percentile of clinically diagnosed AD dementia [16]. We adopted this last approach in 2012 [16] for amyloid PET, FDG PET, and structural MRI with the assumption that the same method should be used to select cut points for all biomarkers. However, we now believe that it may be appropriate to select cut points for different AD biomarkers using different methods. In particular, it seems reasonable to treat amyloid biomarkers differently from others. Defining cut points using individuals that meet certain clinical criteria without regard to evidence of amyloidosis is problematic [17]. The field has reached a consensus that biomarker evidence of amyloidosis is necessary for an accurate diagnosis of AD in living persons [3,4,6]. Of individuals with clinically diagnosed AD dementia, 10%–30% do not have AD at autopsy [18] or have no biomarker evidence of amyloidosis [19,20]. Therefore, using a clinical diagnosis of AD dementia to define an "affected" group of cases with AD when selecting biomarker cut points has significant inherent error. Similarly, around 30% of clinically normal elderly individuals have AD at autopsy [21] or have biomarker evidence of amyloidosis [22–24], and therefore, a clinically defined "unaffected" non-AD control group also has significant inherent error [17].

Tau PET has recently been introduced [25–34], and defining a normal/abnormal cut point is needed to place this modality on the same footing with other AD biomarkers. This in turn provides an opportunity to revisit the issue of defining cut points for more established imaging biomarkers used in AD research. Our objective was to examine different methods for defining cut points for amyloid PET, structural MRI, FDG PET, and tau PET. Identifying a single "best" cut point for each biomarker would provide the most straightforward outcome. However, "best" depends on the context of use [35], and therefore, it is reasonable that different cut points might apply for a given biomarker when used for different purposes [36].

In practice, biomarkers vary in terms of whether numerically high or low values are more abnormal. To simplify our presentation, we have reversed the axes for FDG PET and cortical thickness so that from left to right or bottom to top values are increasingly abnormal. In our general discussion of biomarkers, we treat higher values as more abnormal.

2. Methods

2.1. Participants

All clinically normal (CN) individuals in this study were participants enrolled in the Mayo Clinic Study of Aging (MCSA) [37]. Individuals with mild cognitive impairment (MCI) or AD dementia were participants enrolled in either the MCSA or the Mayo Alzheimer's Disease Research Center (ADRC). Beginning in 2004, the MCSA enrolled individuals aged 70–89 years; in 2012, the MCSA began enrolling individuals 50+ years; and, in 2015, the MCSA began enrolling individuals 30+ years. From 2006 to 2015, the imaging battery consisted of MRI, FDG PET, and amyloid PET. In 2015, tau PET was added to this battery, and FDG PET became optional.

All individuals included in this study completed MRI and amyloid PET imaging. However, owing to changes in the MCSA enrollment protocol, not all completed tau PET and FDG PET. Because of its recent introduction, only 508 individuals have tau PET scans. To take advantage of all available data, we created two separate samples for our analyses. The first sample included all individuals with tau PET, amyloid PET, and MRI (many of whom also had FDG PET). We refer to this sample as the "tau/amyloid/MRI sample." The second sample included all individuals with amyloid PET, FDG PET, and MRI. We refer to this sample as the "amyloid/FDG/MRI sample." Some of these individuals also had tau PET imaging. If individuals had multiple imaging visits, we used the first available visit with the necessary modalities.

The evolution of the MCSA described above has several practical implications. First, there are relatively few individuals under age 50 years. As the start of tau PET scanning coincided with enrolling this younger age group, all who consented to imaging had tau PET, amyloid PET, and MRI. Second, serial imaging data are only available in individuals age 50 years or older and only available for amyloid PET, FDG PET, and MRI.

2.2. Standard protocol approvals, registrations, and patient consents

These studies were approved by the Mayo Clinic and Olmsted Medical Center institutional review boards and written informed consent was obtained from all participants.

2.3. Experimental design

2.3.1. Imaging methods

Amyloid PET imaging was performed with Pittsburgh Compound B [38] and tau PET with AV1451 [29]. Computed tomography was obtained for attenuation correction. Late uptake amyloid PET images were acquired from 40–60 minutes, FDG from 30–40 minutes, and tau PET from 80–100 minutes after injection. PET images were analyzed with our in-house fully automated image processing pipeline [39], where image voxel values are extracted from automatically labeled regions of interest (ROIs) propagated from an MRI template. An amyloid PET standardized uptake value ratio (SUVR) was formed from the voxel-number weighted average of the median uptake in the prefrontal, orbitofrontal, parietal, temporal, anterior and posterior cingulate, and precuneus ROIs normalized to the cerebellar crus gray median. Amyloid PET values are expressed both in SUVR units and in centiloid units. The SUVR to centiloid conversion was done as recommended in Klunk et al [40]. An AD-signature FDG PET composite

or "meta-ROI" was formed from the voxel-number weighted average of the median uptake in the angular gyrus, posterior cingulate, and inferior temporal cortical ROIs and normalized to the pons and vermis median [41]. A tau PET meta-ROI was formed from a voxel-number weighted average of the median uptake in the entorhinal, amygdala, parahippocampal, fusiform, inferior temporal, and middle temporal ROIs normalized to the cerebellar crus gray median. PET data were not partial volume corrected. However, the data were "sharpened"—that is, voxels whose probability of being cerebrospinal fluid (CSF) was greater than the probability of being gray matter and greater than the probability of being white matter were discarded from all PET ROI measures.

MRI was performed on one of three 3T GE systems. The MRI measure was a FreeSurfer (v5.3)-derived AD-signature meta-ROI composed of the surface-area weighted average of the mean cortical thickness in the following individual ROIs: entorhinal, inferior temporal, middle temporal, and fusiform.

2.4. Statistical analysis

2.4.1. Methods of defining cut points

We used five methods for determining biomarker cut points which we term (1) reliable worsening, (2) specificity, (3) sensitivity, (4) accuracy of cognitively impaired versus younger CN, and (5) accuracy of cognitively impaired versus age-matched CN. We summarize the five methods graphically in Fig. 1 and describe the methodology in detail below.

The reliable worsening cut point was based on identifying a threshold baseline value beyond which the rate of change in that biomarker worsens reliably. With this method, we first estimated the annual rate of change in each biomarker within each individual with serial imaging data using linear regression. Next, a nonparametric scatter plot smoother was used to plot the mean rate of change in biomarker on the y-axis versus the baseline biomarker value on the x-axis [42]. We then identified the biomarker value that corresponded to the minimum point on this rate versus baseline curve. Although the minimum corresponds to a threshold beyond which the biomarker rate of change can be expected to increase on average, we obtained a more conservative, or reliable, cut point by projecting the upper bound of a 50% prediction interval at the minimum rightward until it intersected the rate versus baseline curve. The point of this intersection defined a reliable worsening cut point. Because we do not yet have longitudinal tau PET data, this reliable worsening analysis was only applicable to MRI, amyloid, and FDG PET and, as shown in results, produced a result only for amyloid PET. Individuals used in this analysis included all MCSA participants (CN, MCI, and AD dementia) with at least one follow-up imaging study.

The specificity-based cut point corresponded to the 95th percentile of the biomarker distribution among MCSA CN individuals' aged 30–49 years, all of whom had normal amyloid

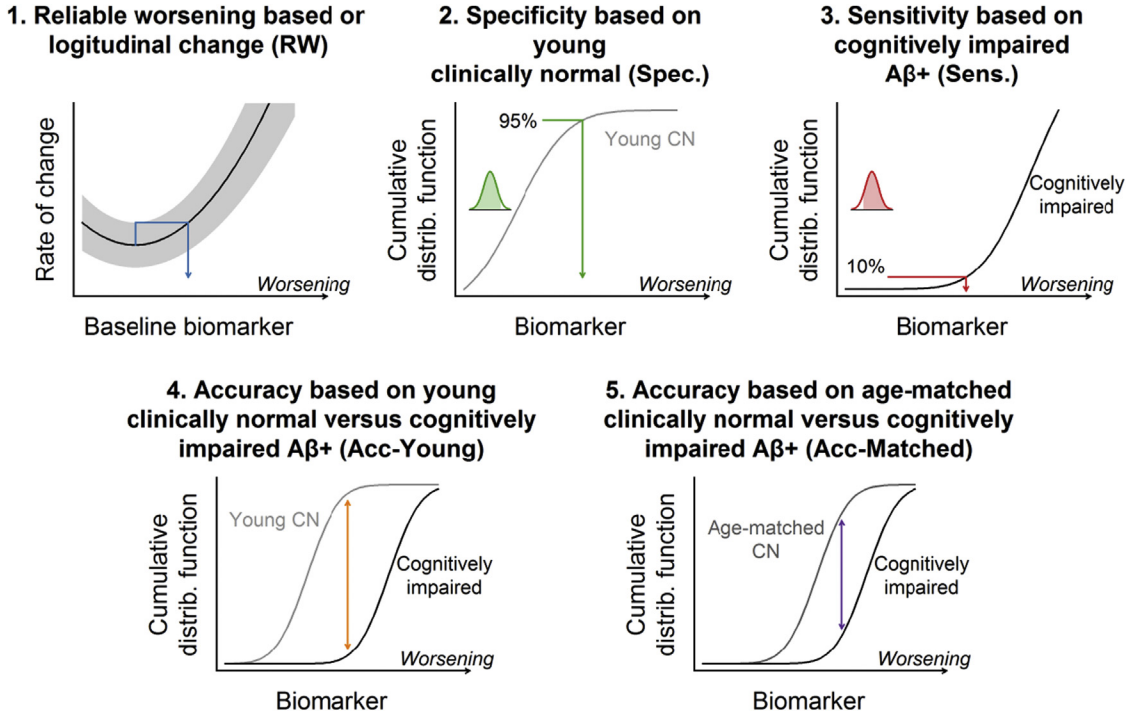

Fig. 1. Cut point methods. Graphical summary of the five methods used for determining cut points. In each panel, increasing numeric values of the biomarker correspond to biomarker worsening. A biomarker's cumulative distribution function (CDF) indicates a biomarker value x on the horizontal axis and the proportion of observations $\leq x$ on the vertical axis.

based on the reliable worsening cut point described previously. As noted previously, we treat all biomarkers such that the 95th percentile corresponds to more abnormal values. The rationale for a specificity-based cut point is that young individuals are likely to be relatively free of AD pathology. This approach is commonly used in laboratory medicine [43]. This approach was applicable to all four imaging methods. Because of limited sample sizes among those with tau PET, we estimated the 95th percentile based on calculating a smoothed cumulative distribution function (CDF) from a kernel density estimate of the distribution. This can be interpreted as a smoothed empirical CDF.

The sensitivity-based cut point corresponded to the 10th percentile of the biomarker distribution among cognitively impaired (aMCI or AD) individuals from the MCSA or ADRC aged 60 years or older who had abnormal amyloid, where the definition of abnormal amyloid was based on the reliable worsening cut point described above. The 10th percentile (i.e., 90% sensitivity) in this impaired group was calculated using the same CDF approach described above. This method is very similar to our earlier approach from 2012 [16]. However, this updated method only included people with abnormal amyloid so that the cognitively impaired group was not contaminated with individuals who are not in the AD pathway [17]. This sensitivity-based cut point is only applicable to selecting the cut point for MRI, FDG PET, and tau PET because the impaired group was required to have elevated values of amyloid PET.

The fourth method was based on discriminating between the cognitively impaired individuals used in the sensitivity cut point analysis and younger CN individuals used in the specificity cut point analysis. This case/control discrimination cut point was obtained by identifying the point of maximum separation between the smoothed CDFs of the two groups. This is equivalent to maximizing accuracy, defined as sensitivity − (1 − specificity). This method is only applicable to selecting cut points for MRI, FDG PET, and tau PET because the cognitively impaired individuals were required to have elevated amyloidosis.

The fifth method was also based on discriminating between cognitively impaired and CN individuals. However, in contrast to the fourth method, the control group consisted of older CN individuals with normal amyloid from the MCSA who were age and sex matched to the impaired group. Using a control group with similar ages as the cognitively impaired group allows this cut point to focus on AD-related differences between the groups as opposed to both AD-specific and nonspecific age-related differences. Requiring the CN individuals to have normal amyloid insures that controls are not in the AD pathway. This method is only applicable to selecting cut points for MRI, FDG PET, and tau PET because amyloid PET was used to define both groups.

2.4.2. *Methods for evaluating cut points*

To evaluate the utility of cut points defined from the different methods above, we estimated the proportion of MCSA CN individuals with abnormal biomarker values by age using logistic regression models. Age was fit with a restricted cubic spline with knots at ages 50, 65, and 80 years. We also generated histograms of the biomarkers among all MCSA individuals aged 50–89 years to illustrate where the cut points fall in the biomarker distributions within a population. These histograms were weighted to reflect the age/sex frequencies in Olmsted County.

3. Results

Characteristics of the clinical groups used are found in Table 1, where we distinguish between the two samples in our study. Supplementary Fig. 1 illustrates how the clinical groups within the tau/amyloid/MRI sample compare to the clinical groups in the amyloid/FDG/MRI sample.

Fig. 2 shows the relationship between the annual rate of change and baseline biomarker for amyloid PET, FDG PET, and cortical thickness among all MCSA individuals with serial imaging. Tau PET is not shown because we only have cross-sectional data currently. Using this reliable worsening method, we defined the amyloid PET cut point as 1.42 SUVR, corresponding to a centiloid value of 19. A reliable worsening cut point could not be determined for FDG or cortical thickness because the rate of change was not significantly associated with the baseline biomarker values.

Fig. 3 shows the estimated cumulative distribution function among younger CN individuals, cognitively impaired individuals, and age-matched and sex-matched CN individuals. The specificity, sensitivity, and accuracy of impaired versus younger CN methods gave very similar cut points within each biomarker: 1.21, 1.20, and 1.23 SUVR for tau PET, respectively; 1.53, 1.56, and 1.56 SUVR for FDG PET; and 2.69, 2.70, and 2.67 mm for cortical thickness. However, cut points defined using the accuracy of impaired versus age-matched CN method were always more conservative (1.33 SUVR for tau PET, 1.41 SUVR for FDG PET, and 2.57 mm for cortical thickness). Although a cut point for amyloid PET could not be determined using the sensitivity or accuracy methods due to circularity in group definitions, the specificity method produced a cut point of 1.30 SUVR, centiloid 8.

In Fig. 4, we illustrate the continuous distribution of amyloid PET, tau PET, FDG PET, and MRI cortical thickness values versus age among all MCSA CN with horizontal lines representing the cut points from each method. A box plot of values among impaired individuals is included in the plot of each biomarker for reference.

Fig. 4 also shows the estimated proportion of CN that are labeled abnormal by age for each of the five cut point methods. For tau PET, the cut point from the sensitivity method labeled the most people abnormal, whereas the specificity and accuracy of impaired versus young CN methods labeled about 10% and 15% fewer people abnormal, respectively. For FDG PET, the sensitivity and accuracy of impaired versus young CN methods (which produced the same cut point) labeled the most people abnormal, whereas the specificity

Table 1
Characteristics of participant samples used in analyses

	Tau/amyloid/MRI sample					Amyloid/FDG/MRI sample				
	Cut point definition samples			Cut point evaluation samples		Cut point definition samples			Cut point evaluation samples	
Characteristic	Young CN (n = 49)	Matched CN (n = 102)*	Cognitively impaired (n = 51)†	All CN (n = 438)	MCSA 50–89 (n = 417)‡	Matched CN (n = 214)*	Cognitively impaired (n = 214)†	MCSA serial (n = 682)§	All CN (n = 1503)	MCSA 50–89 (n = 1646)‡
Age, years										
Median (IQR)	39 (34–44)	73 (67–78)	74 (67–79)	69 (59–78)	71 (64–79)	77 (72–84)	77 (72–84)	75 (68–80)	71 (63–78)	72 (64–79)
Min, max	30–48	60–90	60–88	30–94	52–89	60–93	60–93	51–94	30–95	50–89
Male gender, no. (%)	28 (57%)	70 (69%)	35 (69%)	241 (55%)	236 (57%)	124 (58%)	124 (58%)	399 (59%)	790 (53%)	882 (54%)
Education, years, Median (IQR)	16 (14–16)	14 (13–17)	16 (12–16)	16 (13–17)	15 (13–16)	14 (12–17)	15 (12–16)	14 (12–16)	15 (12–16)	14 (12–16)
APOE ε4 positive, no. (%)	5 (19%)	15 (15%)	33 (72%)	106 (26%)	109 (27%)	29 (14%)	143 (69%)	187 (28%)	392 (27%)	462 (29%)
MMSE, median (IQR)	29 (29–30)	29 (28–29)	24 (22–27)	29 (28–29)	29 (28–29)	28 (27–29)	25 (22–27)	28 (27–29)	29 (28–29)	29 (27–29)
Tau PET, SUVR, median (IQR)	1.12 (1.06–1.14)	1.18 (1.11–1.22)	1.72 (1.37–2.07)	1.18 (1.12–1.23)	1.19 (1.13–1.25)					
Amyloid PET, SUVR, median (IQR)	1.21 (1.17–1.24)	1.32 (1.28–1.37)	2.50 (2.24–2.70)	1.35 (1.27–1.48)	1.36 (1.29–1.53)	1.32 (1.29–1.37)	2.33 (2.00–2.59)	1.36 (1.30–1.54)	1.34 (1.28–1.45)	1.35 (1.29–1.51)
FDG PET, SUVR, median (IQR)	1.74 (1.65–1.82)	1.54 (1.46–1.64)	1.20 (1.12–1.34)	1.57 (1.47–1.66)	1.54 (1.45–1.64)	1.54 (1.44–1.64)	1.32 (1.23–1.44)	1.55 (1.45–1.65)	1.58 (1.47–1.68)	1.56 (1.45–1.66)
Cortical thickness, mm, median (IQR)	2.86 (2.78–2.93)	2.70 (2.60–2.79)	2.39 (2.27–2.60)	2.73 (2.63–2.83)	2.70 (2.61–2.80)	2.65 (2.58–2.73)	2.50 (2.35–2.60)	2.67 (2.59–2.77)	2.71 (2.62–2.81)	2.70 (2.60–2.80)

*The matched CN subgroups include clinically normal individuals from the MCSA aged 60 or older with amyloid PET < 1.42 SUVR and are matched on age (within 3 years) and sex to the cognitively impaired subgroups. The tau/amyloid/MRI matched CN sample was matched 2 to 1 to the tau/amyloid/MRI cognitively impaired sample. The amyloid/FDG/MRI matched CN sample was matched 1 to 1 to the amyloid/FDG/MRI cognitively impaired sample.

†The tau/amyloid/MRI cognitively impaired subgroup includes 19 aMCI and 32 AD dementia individuals with amyloid PET >1.42 SUVR and aged 60 or older. The amyloid/FDG/MRI cognitively impaired subgroup includes 136 aMCI and 78 AD dementia individuals with amyloid PET >1.42 SUVR and aged 60 or older.

‡The tau/amyloid/MRI MCSA 50–89 subgroup includes 384 CN, 27 MCI, 5 individuals with dementia, and 1 individual with an other diagnosis. The amyloid/FDG/MRI MCSA 50–89 subgroup includes 1446 CN, 186 MCI, 12 individuals with dementia, and 2 individuals with an other diagnosis.

§The amyloid/FDG/MRI serial imaging subgroup includes 606 CN, 75 MCI, and 1 individual with dementia.

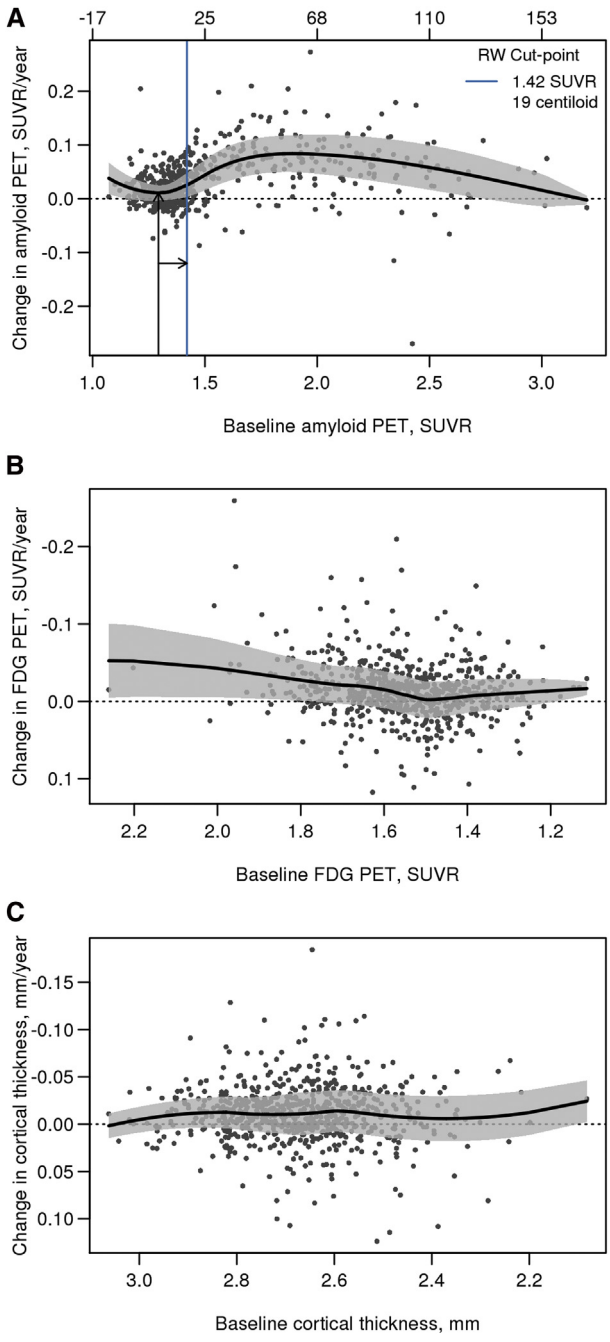

Fig. 2. Reliable worsening cut point. Scatter plot of annual rate of change in imaging biomarker versus baseline with a nonparametric scatter plot smoother line and a 50% prediction interval. For amyloid PET (A), the arrows and solid blue line illustrate how the reliable worsening (RW) cut point was obtained. For this panel, values are shown in both SUVR and centiloid units. A reliable worsening cut point was not obtained for FDG PET or MRI cortical thickness (B and C).

method labeled about 10% fewer people abnormal. Performance was similar across the three methods for cortical thickness, although the sensitivity method labeled more people abnormal and the accuracy of impaired versus young CN fewer by roughly 10%. In contrast, cut points defined using the accuracy of impaired versus age-matched CN method labeled the fewest people abnormal for all biomarkers. For example, at age 70 years, the proportion of CN-labeled abnormal using the specificity, sensitivity, accuracy of impaired versus young CN, and accuracy of impaired versus age-matched CN methods for tau PET is 33%, 42%, 26%, and 5%, respectively. For FDG PET, the proportions were 35%, 45%, 45%, and 9%, and for MRI cortical thickness, the proportions were 37%, 42%, 31%, and 10%, respectively. For amyloid PET, 30% of CN individuals are labeled abnormal at age 70 using the reliable worsening cut point, in contrast to 75% using the specificity cut point.

In Fig. 5, we illustrate the histograms of the distribution of values among all MCSA individuals ages 50–89 years along with the cut points selected by the different methods for each. The histograms have been weighted by age and sex to reflect the age/sex population frequencies in Olmsted county. We also show the distribution of a prototypical non-imaging and non-AD biomarker, systolic blood pressure, to illustrate how an established biomarker and its accepted cut point compares with those used in AD research. MRI, FDG PET, tau PET, and systolic blood pressure are all approximately normally distributed. The frequency distribution of amyloid PET is unimodal but has a prominent skew.

Supplementary Fig. 2 illustrates how the proportion of MCSA CN individuals labeled abnormal by age using the cut point definitions for amyloid PET, FDG PET, and MRI cortical thickness defined in this analysis compare to our previous cut point definitions [44]. Due to changes in the implementations of the biomarker processing pipelines the exact numeric values reported in this paper are not directly comparable to previously reported values.

4. Discussion

4.1. Cut point selection for amyloid PET

Of the five cut point methods considered, only the reliable worsening and specificity methods are applicable to amyloid PET, because amyloid was used in the group definitions for the other methods. However, the specificity cut point of SUVR 1.30, centiloid 8, classifies 75% of CN individuals in the MCSA at age 70 years as above the amyloid threshold. A cut point indicating most 70-year olds have abnormal amyloid is inconsistent with autopsy data which indicate that the proportion of the population with amyloidosis at age 70 years is just under 40% [45]. In vivo amyloid imaging significantly underestimates actual b-amyloid load at autopsy (Colin Masters, Alzheimer's Association International Conference, 2016); therefore, the cut point obtained from the sensitivity method is not plausible for amyloid PET. One possible, although speculative, explanation is illustrated by Fig. 4 where values among individuals with low-amyloid PET uptake tend to steadily increase with age. We suspect that nonspecific ligand uptake in the white matter might increase with age. Nonspecific white matter uptake would bleed into cortical target ROIs. The result would be that

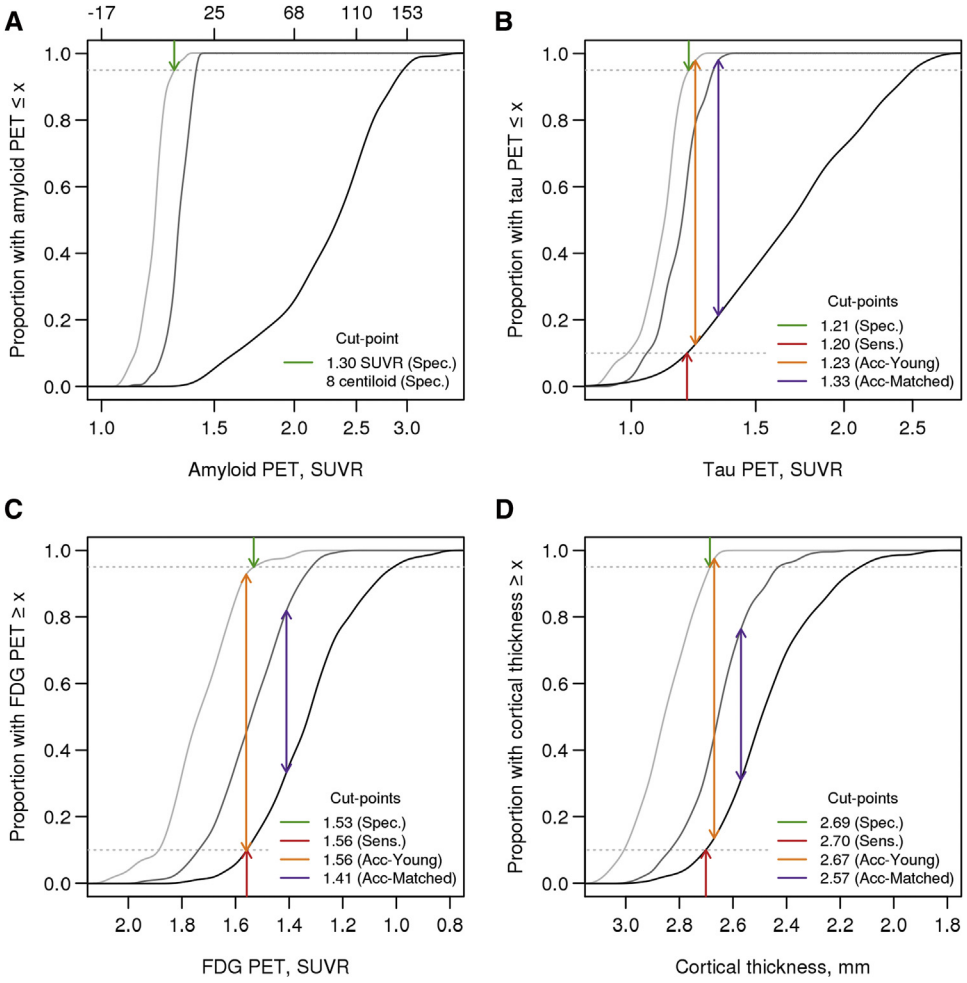

Fig. 3. Specificity, sensitivity, and accuracy cut points. Cumulative distribution function (CDF) plots for young CN individuals (light gray), cognitively impaired individuals (black), and older CN individuals that were age and sex matched to the cognitively impaired group (dark gray). The arrows indicate cut points chosen corresponding to 95% specificity (CDF = 0.95, dark green arrow), 90% sensitivity (CDF = 0.10, dark red arrow), accuracy in discriminating between young CN and cognitively impaired individuals (orange arrow), and accuracy in discriminating between age-matched CN and cognitively impaired individuals (purple arrow). Accuracy was defined as the point of maximum difference between two CDFs. Amyloid PET was used in the definition of the cognitively impaired group so only the specificity cut point is shown. For amyloid PET (A), values are shown in both SUVR and centiloid units. (B) tau PET; (C) FDG PET; and (D) MRI.

when a cut point for amyloid PET based on young people is applied to elderly normals, an unrealistically high proportion is labeled abnormal.

Selecting a cut point based on a threshold value beyond which the rate of change in that biomarker worsens reliably has high face validity. It is a particularly useful approach for amyloid because the properties of the biomarker itself are used to select the cut point, independent from any relationship to clinical symptoms [46]. The amyloid PET cut point selected by this reliable worsening method of 1.42 SUVR is similar to the 1.40 SUVR we used in the past, although we made some changes to the implementation of the amyloid PET quantitative pipeline. These include no partial volume correction and a slightly different cerebellar reference ROI. Nonetheless, our old and new amyloid cut points labeled very similar proportions of the MCSA population as abnormal across all ages (Supplementary Fig. 2). Using methods described in Murray et al. [14], with our new cut point of 1.42 5/7 (71%) of autopsied individuals with Thal phase < 2 are labeled normal, and 23/25 (92%) with Thal phase ≥ 2 are labeled abnormal. Thus, 1.42 SUVR with our new method seems to be a well determined cut point for amyloid PET.

SUVR values are laboratory specific units that depend on the tracer used and the methodological implementation of data processing. The centiloid concept was introduced for amyloid PET to enable the field to express quantitative amyloid PET in universal units [40]. A cut point of 1.42 SUVR with our current data processing methods corresponds to a centiloid value of 19.

4.2. Approach to tau PET quantification

We selected the entorhinal, amygdala, parahippocampal, fusiform, inferior temporal, and middle temporal ROIs for

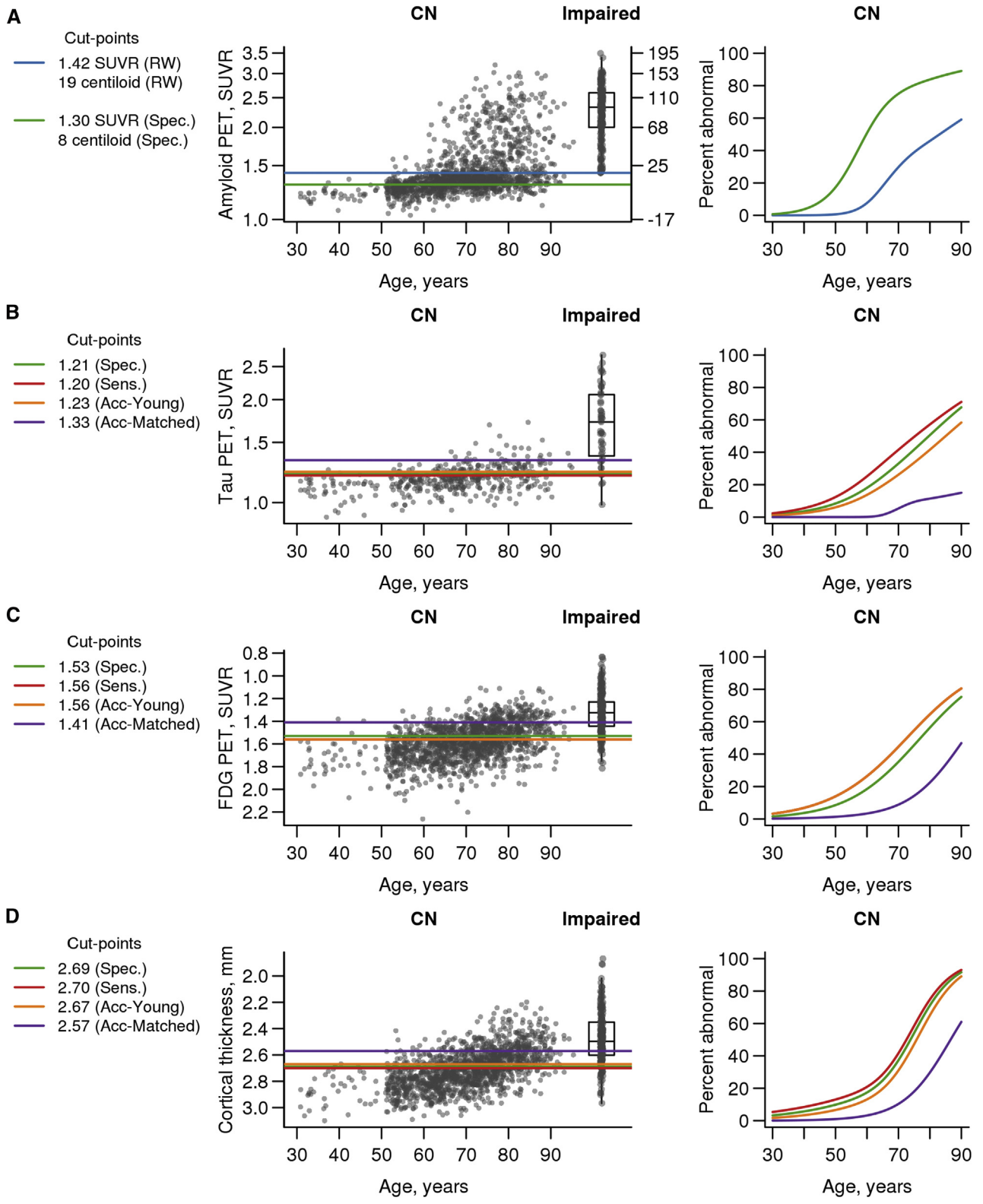

Fig. 4. Biomarkers versus age. Scatter plots of each biomarker versus age among all Mayo Clinic Study of Aging (MCSA) CN individuals with box plots of the cognitively impaired group shown for reference. The horizontal lines indicate cut points chosen from the five methods. The colors used were as follows: reliable worsening (RW), blue (only in [A]); specificity (Spec.), green; sensitivity (Sens.), red; accuracy of cognitively impaired versus young CN (Acc-Young), orange; accuracy of cognitively impaired versus matched CN (Acc-Matched), purple. Using each of these five cut point methods, we then label individuals as normal or abnormal and show the percent abnormal by age as estimated from logistic regression models. For amyloid PET (A), values are shown in both SUVR and centiloid units. (B) tau PET; (C) FDG PET; and (D) MRI.

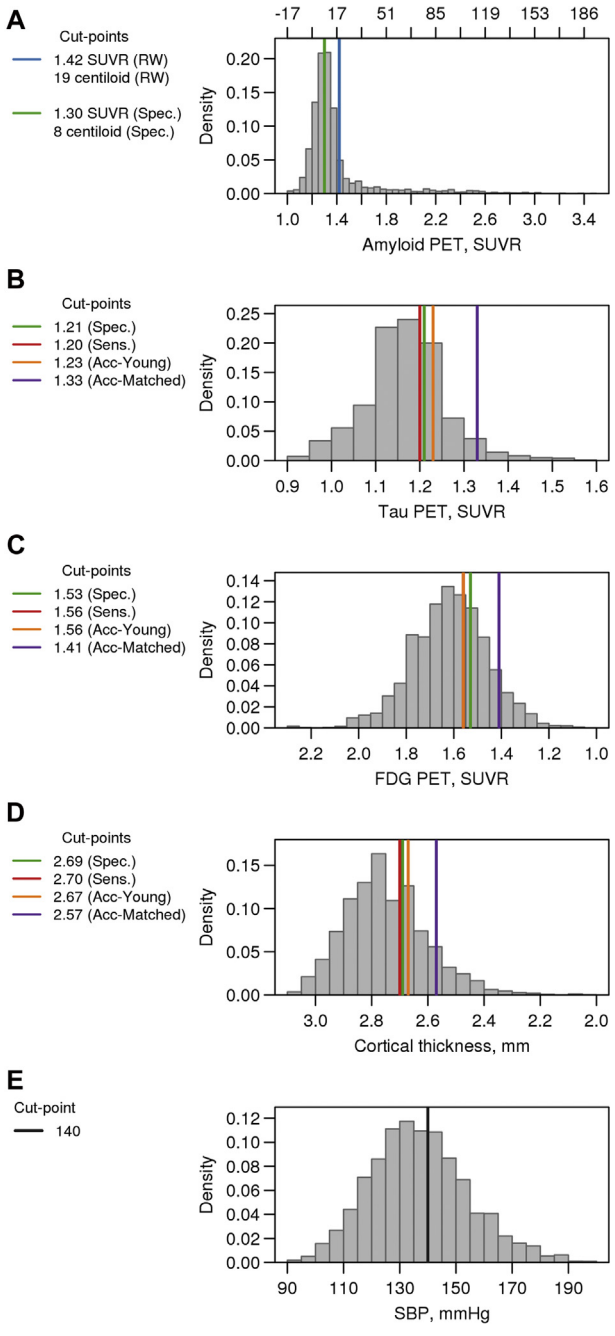

Fig. 5. Distribution of biomarkers in the population. Histograms of the distribution of each biomarker (A–D) among all Mayo Clinic Study of Aging (MCSA) individuals aged 50–89 years weighted to the Olmsted county population by age and sex. The vertical lines indicate cut points chosen from the five methods. The colors used were as follows: reliable worsening (RW), blue ([A] only); specificity (Spec.), green; sensitivity (Sens.), red; accuracy of cognitively impaired versus young CN (Acc-Young), orange; accuracy of cognitively impaired versus matched CN (Acc-Matched), purple. Systolic blood pressure is also shown in (E) with the cut point of 140 mm Hg (black line). For amyloid PET (A), values are shown in both SUVR and centiloid units.

our tau meta-ROI because ligand uptake in medial temporal ROIs is characteristic of clinically normal individuals with advancing age [27–29,31–34]. Ligand uptake in basal/lateral temporal lobe—fusiform, inferior temporal, and middle temporal—is associated with characteristics of AD such as abnormal amyloid, worse cognitive performance in CN individuals, and the clinical diagnosis of MCI or AD dementia [27–34]. Temporal lobe tau PET uptake also predicts elevated (abnormal) CSF tau [30,33]. Thus, this composite set of ROIs captures a broad dynamic range across the normal to pathologic aging to AD dementia spectrum [30]. We did not include the hippocampus in the tau PET meta-ROI because of frequent bleed-in from off-target uptake in the choroid plexus.

An argument can be made that tau PET should be reported in stages analogously to Braak stage [29,34,47] rather than using a dichotomous normal/abnormal designation. However, a tau PET quantification method based on the magnitude of uptake in an AD-signature composite-ROI that captures a broad diagnostic range, as reported in this study, encapsulates the concept of pathological severity. Furthermore, most tests used in medicine have some notion of a normal/abnormal cut point in their continuous distributions, and an effort to accomplish the same for tau PET seems logical.

4.3. Selecting cut points for tau PET, FDG PET, and MRI

For MRI and FDG PET, there was no clear evidence of a threshold value beyond which rates of hypometabolism or cortical thinning increase. Thus, the reliable worsening method did not produce a cut point for FDG PET or MRI (Fig. 2). As noted above, with no serial data, a reliable worsening cut point for tau PET could not be obtained. Using the other cut point methods of specificity, sensitivity, and accuracy of cognitively impaired versus young CN produced more lenient cut points, whereas the accuracy of cognitively impaired versus age-matched CN method produced more conservative cut points.

The fact that several of the cut point methods produced very consistent results for tau PET, FDG PET, and MRI lends support to the validity of these cut points. We emphasize that regardless of how lenient or conservative a cut point is, if an individual falls below that cut point for a particular biomarker, this does not imply that no pathology is present in the brain. An individual labeled normal may well have pathology in the brain but not at a sufficient level to cross the in vivo detection threshold of the biomarker in question.

The amyloid PET cut points are reported in standardized centiloid units [40], and the MRI thickness cut points are reported in mm using standard freely available FreeSurfer methods and thus are usable by other groups. This is not the case for tau PET and FDG PET cut points because the specific SUVR values depend on our own imaging processing pipelines. However, for all four modalities, there is value in comparing results of different methods of selecting cut points and examining the consequences of applying different cut points to a clinically normal sample of individuals that spans a large age range. Different research groups can

examine these results and use them to guide selection of cut point methods that best suit their own research needs. For example, cut points from the accuracy of impaired versus age-matched CN method might be used if the research question centered on whether individuals were abnormal relative to what is expected for age (with the understanding that AD biomarkers are abnormal in many clinically normal elderly). In contrast, cut points from the accuracy of impaired versus young CN method might be used if the research question centered on whether individuals were abnormal in an absolute sense without accounting for aging effects. The choice of method could also depend on whether one wants to capture the earliest stages of disease versus wanting to identify people who are clearly advanced on the AD pathway.

4.4. Dispelling myths about cut points

Examination of the data in Fig. 5 should challenge several popular assumptions about AD biomarkers. The first is that amyloid PET (and CSF Ab42) is bimodally distributed in "the population" [48]. While that is true in highly selected samples composed mostly of impaired individuals [48], it is not true in our population-based sample where most are not demented (Fig. 5). A second assumption is that cut points for any AD biomarker other than amyloid are not valid because these biomarkers are not bimodally distributed. The distribution of systolic blood pressure in Fig. 5 illustrates an example of one of many established biomarkers that is approximately normally distributed and has a cut point that is widely used clinically.

4.5. Group-wise separation

Although our main focus was examining different cut point methods and the consequences of these different approaches, Fig. 4 does reveal that the overlap in tau PET values between elderly CN (over 80) and the impaired group is considerably less than the overlap present between these two groups for both FDG PET and MRI. One interpretation of this is that hypometabolism and atrophy reflect the cumulative effects of brain damage from many possible etiologies experienced by an individual. In any individual, brain damage may be the product of rapidly progressive active neurodegenerative processes, slowly progressive active neurodegenerative processes, past insults that are now static, or combinations of these. Aging alone in the absence of obvious pathology is associated with loss of synapses and dendrites, loss of function, and neuronal shrinkage [49]. All elderly individuals have experienced the effects of aging on brain structure and metabolism, and many have experienced or are experiencing other insults as well which accounts for overlap in these measures between clinically normal elderly and cognitively impaired individuals. In contrast, tau PET seems to largely index AD-like tau. Therefore, high-tau levels (especially isocortical tau) appear to be less compatible with preserved cognition than either low cortical thickness or hypometabolism.

4.6. Methodological considerations and limitations

We conditioned the samples used to define many of the cut points for tau PET, FDG PET, and MRI on amyloid because our focus was AD centric. This AD centricity guided our a priori selection of meta-ROIs to include regions that are characteristically involved in AD: medial-basal temporal atrophy for MRI, medial, basal, and lateral temporal lobe for tau, and temporal-parietal hypometabolism for FDG. However, the fact that these brain areas are characteristic of AD does not make these imaging signatures specific for AD. Non-AD pathologies like alpha synuclein, TDP 43, agyrophillic grains, micro infarctions, and so forth can also result in medial-basal temporal atrophy or temporal-parietal hypometabolism. Therefore, when applying these cut points to a general population, individuals both on AD as well as non-AD–related pathways may be classified as abnormal on these markers. We have previously labeled these non-AD related abnormal individuals as suspected non-Alzheimer's pathophysiology [16], an important group that differs in clinical and cognitive outcomes from those individuals in the AD pathway.

A limitation of our study was the relatively small number of young individuals (n = 49) and cognitively impaired individuals (n = 51). To place this in context, however, the centiloid data set [40] which is the standard by which all groups are indexing amyloid PET values consists of 34 young controls and 45 AD dementia patients.

Hippocampal volume is widely used in the field; however, we did not include this measure in our analyses. We have recently moved away from hippocampal volume as an index of cumulative neuronal damage and instead now favor the AD-signature cortical thickness measure, we used in the article [44]. The primary reason for this is that surface area and hence also brain volume measures scale with head size necessitating an adjustment [50]. Correcting for this is not straightforward because (1) head size is related to sex, (2) the premorbid relationship between volume and head size (which cannot be observed in subjects older than ~50) probably differs by sex, and (3) there are likely to be important sex-specific causes of atrophy. Although we hope to investigate how to optimally account for the confounding without introducing artifactual differences due to over adjustment, a very straightforward solution is to use a measure of cortical thickness which does not scale closely with head size and does not require any adjustment.

Acknowledgments

This study was supported by the National Institute of Health (R01 AG011378, R01 AG041851, U01 AG006786, and R01 AG034676), the GHR Foundation, and the Alexander

Family Alzheimer's Disease Research Professorship of the Mayo Foundation.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2016.08.005.

RESEARCH IN CONTEXT

1. Systematic review: The authors reviewed the literature using traditional (e.g., PubMed) sources and meeting abstracts and presentations.
2. Interpretation: We have identified a single cut point for amyloid PET (SUVR 1.42, centiloid 19) and both lenient and conservative cut points for tau PET, FDG PET and MRI based on principled approaches. To optimize consistency in approach, our future plan is to base lenient cut points for tau PET, FDG PET, and cortical thickness on the cognitively impaired versus young CN method (1.23 SUVR for tau PET, 1.56 SUVR for FDG PET; and 2.67 mm for cortical thickness) and to base conservative cut points on the cognitively impaired versus age-matched CN method (1.33 SUVR for tau PET, 1.41 SUVR for FDG PET, and 2.57 mm for cortical thickness).
3. Future directions: We outline a principled approach to selecting biomarker cut points that can aid the development of diagnostic criteria that hinge on biomarker results.

References

[1] Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med 2014;6:228fs13.
[2] Hampel H, Schneider LS, Giacobini E, Kivipelto M, Sindi S, Dubois B, et al. Advances in the therapy of Alzheimer's disease: targeting amyloid beta and tau and perspectives for the future. Expert Rev Neurother 2015;15:83–105.
[3] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 2014;13:614–29.
[4] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and Alzheimer's Association Workgroup. Alzheimers Dement 2011;7:270–9.
[5] Jack CR Jr, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carillo M, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:257–62.
[6] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging and the Alzheimer's Association Workgroup. Alzheimers Dement 2011;7:263–9.
[7] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–92.
[8] Mormino EC, Brandel MG, Madison CM, Rabinovici GD, Marks S, Baker SL, et al. Not quite PIB-positive, not quite PIB-negative: Slight PIB elevations in elderly normal control subjects are biologically relevant. Neuroimage 2012;59:1152–60.
[9] Villeneuve S, Rabinovici GD, Cohn-Sheehy BI, Madison C, Ayakta N, Ghosh PM, et al. Existing Pittsburgh Compound-B positron emission tomography thresholds are too high: statistical and pathological evaluation. Brain 2015;138:2020–33.
[10] Pike KE, Savage G, Villemagne VL, Ng S, Moss SA, Maruff P, et al. Beta-amyloid imaging and memory in non-demented individuals: evidence for preclinical Alzheimer's disease. Brain 2007;130:2837–44.
[11] Rowe CC, Ellis KA, Rimajova M, Bourgeat P, Pike KE, Jones G, et al. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol Aging 2010;31:1275–83.
[12] Hedden T, Van Dijk KR, Becker JA, Mehta A, Sperling RA, Johnson KA, et al. Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. J Neurosci 2009;29:12686–94.
[13] Villemagne VL, Pike KE, Chetelat G, Ellis KA, Mulligan RS, Bourgeat P, et al. Longitudinal assessment of Abeta and cognition in aging and Alzheimer disease. Ann Neurol 2011;69:181–92.
[14] Murray ME, Lowe VJ, Graff-Radford NR, Liesinger AM, Cannon A, Przybelski SA, et al. Clinicopathologic and 11C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum. Brain 2015;138:1370–81.
[15] Thal DR, Beach TG, Zanette M, Heurling K, Chakrabarty A, Ismail A, et al. [(18)F]flutemetamol amyloid positron emission tomography in preclinical and symptomatic Alzheimer's disease: Specific detection of advanced phases of amyloid-beta pathology. Alzheimers Dement 2015;11:975–85.
[16] Jack CR Jr, Knopman DS, Weigand SD, Wiste HJ, vemuri P, Lowe V, et al. An operational approach to NIA-AA crtiteria for preclinical Alzheimer's disease. Ann Neurol 2012;71:765–75.
[17] Wisse LE, Butala N, Das SR, Davatzikos C, Dickerson BC, Vaishnavi SN, et al. Suspected non-AD pathology in mild cognitive impairment. Neurobiol Aging 2015;36:3152–62.
[18] Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J Neuropathol Exp Neurol 2012;71:266–73.
[19] Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. N Engl J Med 2014;370:322–33.
[20] Ossenkoppele R, Jansen WJ, Rabinovici GD, Knol DL, van der Flier WM, van Berckel BN, et al. Prevalence of amyloid pet positivity in dementia syndromes: A meta-analysis. JAMA 2015;313:1939–49.
[21] Knopman DS, Parisi JE, Salviati A, Floriach-Robert M, Boeve BF, Ivnik RJ, et al. Neuropathology of cognitively normal elderly. J Neuropathol Exp Neurol 2003;62:1087–95.
[22] Aizenstein HJ, Nebes RD, Saxton JA, Price JC, Mathis CA, Tsopelas ND, et al. Frequent amyloid deposition without significant cognitive impairment among the elderly. Arch Neurol 2008;65:1509–17.
[23] Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FR, et al. Prevalence of cerebral amyloid pathology in persons without dementia: A meta-analysis. JAMA 2015;313:1924–38.

[24] Mielke MM, Wiste HJ, Weigand SD, Knopman DS, Lowe VJ, Roberts RO, et al. Indicators of amyloid burden in a population-based study of cognitively normal elderly. Neurology 2012;79:1570–7.
[25] Villemagne VL, Fodero-Tavoletti MT, Masters CL, Rowe CC. Tau imaging: early progress and future directions. Lancet Neurol 2015; 14:114–24.
[26] Villemagne VL, Furumoto S, Fodero-Tavoletti MT, Mulligan RS, Hodges J, Harada R, et al. In vivo evaluation of a novel tau imaging tracer for Alzheimer's disease. Eur J Nucl Med Mol Imaging 2014; 41:816–26.
[27] Johnson KA, Shultz A, Betensky RA, Becker JA, Sepulcre J, Rentz DM, et al. Tau positron emission tomographic imaging in aging and early Alzheimer's disease. Ann Neurol 2016;79:110–9.
[28] Scholl M, Lockhart SN, Schonhaut DR, O'Neil JP, Janabi M, Ossenkoppele R, et al. PET Imaging of tau deposition in the aging human brain. Neuron 2016;89:971–82.
[29] Schwarz AJ, Yu P, Miller BB, Shcherbinin S, Dickson J, Navitsky M, et al. Regional profiles of the candidate tau PET ligand 18F-AV-1451 recapitulate key features of Braak histopathological stages. Brain 2016;139:1539–50.
[30] Brier MR, Gordon B, Friedrichsen K, McCarthy J, Stern A, Christensen J, et al. Tau and Aβ imaging, CSF measures, and cognition in Alzheimer's disease. Sci Transl Med 2016;8:338ra66.
[31] Lowe V, Wiste HJ, Pandey M, Senjem M, Boeve B, Josephs KA, et al. Tau-PET imaging with AV-1451 in Alzheimer's disease. In: Johnson KA, Jagust W, Klunk W, Mathis C, eds. Human Amyloid Imaging. Miami Beach, FL: World Events Forum, Inc; 2016. 114.
[32] Ossenkoppele R, Schonhaut DR, Scholl M, Lockhart SN, Ayakta N, Baker SL, et al. Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. Brain 2016;139:1551–67.
[33] Gordon BA, Friedrichsen K, Brier M, Blazey T, Su Y, Christensen J, et al. The relationship between cerebrospinal fluid markers of Alzheimer pathology and positron emission tomography tau imaging. Brain 2016;139:2249–60.
[34] Cho H, Choi JY, Hwang MS, Kim YJ, Lee HM, Lee HS, et al. In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum. Ann Neurol 2016;80:247–58.
[35] Klunk WE, Cohen A, Bi W, Weissfeld L, Aizenstein H, McDade E, et al. Why we need two cutoffs for amyloid-imaging: Early versus Alzheimer's-like amyloid-positivity. Alzheimer's Association International Conference. Vancouver, British Columbia, Canada: Alzheimer's Associaton; 2012. p. P453–P454.
[36] Bartlett JW, Frost C, Mattsson N, Skillback T, Blennow K, Zetterberg H, et al. Determining cut-points for Alzheimer's disease biomarkers: statistical issues, methods and challenges. Biomark Med 2012;6:391–400.
[37] Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, et al. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology 2008;30:58–69.
[38] 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.
[39] Senjem ML, Gunter JL, Shiung MM, Petersen RC, Jack CR Jr. Comparison of different methodological implementations of voxel-based morphometry in neurodegenerative disease. Neuroimage 2005;26:600–8.
[40] Klunk WE, Koeppe RA, Price JC, Benzinger T, Devous M, Jagust W, et al. The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement 2015;11:1–15.
[41] Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. Neurobiol Aging 2011;32:1207–18.
[42] Villain N, Chetelat G, Grassiot B, Bourgeat P, Jones G, Ellis KA, et al. Regional dynamics of amyloid-beta deposition in healthy elderly, mild cognitive impairment and Alzheimer's disease: a voxelwise PiB-PET longitudinal study. Brain 2012;135:2126–39.
[43] Grasbeck R. The evolution of the reference value concept. Clin Chem Lab Med 2004;42:692–7.
[44] Jack CR, Wiste HJ, Weigand S, Knopman D, Mielke MM, Vemuri P, et al. Different definitions of neurodegeneration produce similar frequencies of amyloid and neurodegeneration biomarker groups by age among cognitively non-impaired individuals. Brain 2015; 138:3747–59.
[45] Braak H, Thal DR, Ghebremedhin E, Del Tredici K. Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. J Neuropathol Exp Neurol 2011;70:960–9.
[46] Ingelsson M, Fukumoto H, Newell KL, Growdon JH, Hedley-Whyte ET, Frosch MP, et al. Early Abeta accumulation and progressive synaptic loss, gliosis, and tangle formation in AD brain. Neurology 2004;62:925–31.
[47] Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. Acta Neuropathol 1991;82:239–59.
[48] De Meyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, De Deyn PP, et al. Diagnosis-Independent Alzheimer Disease Biomarker Signature in Cognitively Normal Elderly People. Arch Neurol 2010;67:949–56.
[49] Jagust W. Vulnerable neural systems and the borderland of brain aging and neurodegeneration. Neuron 2013;77:219–34.
[50] Barnes J, Ridgway GR, Bartlett J, Henley SM, Lehmann M, Hobbs N, et al. Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage 2010;53:1244–55.