Published in final edited form as:

Clin Nucl Med. 2017 August ; 42(8): 577–581. doi:10.1097/RLU.0000000000001693.

Augmenting Amyloid PET Interpretations with Quantitative Information Improves Consistency of Early Amyloid Detection

Nicholas R. Harn^a, Suzanne L. Hunt^b, Jacqueline Hill^a, Eric Vidoni^c, Mark Perry^a, and Jeffrey M. Burns^{c,*}

^aRadiology Department, University of Kansas Medical Center

^bBiostatistics Department, University of Kansas Medical Center

^cNeurology Department, University of Kansas Medical Center

Abstract

Purpose—Establishing reliable methods for interpreting elevated cerebral amyloid- β plaque on PET scans is increasingly important for radiologists, as availability of PET imaging in clinical practice increases. We examined a three-step method to detect plaque in cognitively normal older adults, focusing on the additive value of quantitative information during the PET scan interpretation process.

Methods—Fifty-five F-18 florbetapir PET scans were evaluated by three experienced raters. Scans were first visually interpreted as having “elevated” or “non-elevated” plaque burden (“Visual Read”). Images were then processed using a standardized quantitative analysis software (MIMneuro) to generate whole brain and region of interest (ROI) standard uptake value ratios (SUVRs). This “Quantitative Read” was considered elevated if at least two of six ROIs had an SUVR >1.1 . The final interpretation combined both visual and quantitative data together (“VisQ Read”). Cohen’s kappa values were assessed as a measure of interpretation agreement.

Results—Plaque was elevated in 25.5–29.1% of the 165 total Visual Reads. Inter-rater agreement was strong (Kappa=0.73–0.82) and consistent with reported values. Quantitative Reads were elevated in 45.5% of participants. Final VisQ Reads changed from initial Visual Reads in 16 (9.7%) interpretations, with most changing from “non-elevated” Visual Reads to “elevated.” These changed interpretations demonstrated lower plaque quantification than those initially read as “elevated” that remained unchanged. Inter-rater variability improved for VisQ Reads with the addition of quantitative information (Kappa=0.88–0.96).

Conclusion—Inclusion of quantitative information increases consistency of PET scan interpretations for early detection of cerebral amyloid- β plaque accumulation.

Keywords

PET; brain imaging; florbetapir; inter-rater variability; standardized uptake value ratio

*Corresponding Author: Jeffrey M. Burns. 3599 Rainbow Boulevard Kansas City, KS 66160. Ph: (913) 588-0555. Fax: (913) 945-5035. jburns2@kumc.edu.

No conflicts or other disclosures.

INTRODUCTION

Alzheimer's disease (AD) is a growing concern in the U.S. Over five million Americans are currently affected by the disease, with an expected tripling in prevalence by 2050 [1]. While diagnosis is now largely based upon clinical data, the 2011 revision of AD diagnostic criteria calls for the incorporation of biomarkers into the diagnostic algorithm [2]. Additionally, new Positron Emission Tomography (PET) imaging techniques are now capable of detecting AD-relevant molecular changes, including the hallmark pathologies of amyloid- β plaques and neurofibrillary tangles, making it likely these imaging techniques will play an important role in future efforts to detect and diagnose the disease [3]. Three radiopharmaceuticals are currently FDA-approved to detect amyloid- β plaques (florbetapir, flutemetamol, and florbetaben) and new compounds targeting other features of the disease, such as tau, are expected to reach market in the near future [4, 5].

Radiologic interpretation of amyloid- β PET imaging, however, has demonstrated significant inter-reader variability [6]. As such, the initial FDA approval of the first radiotracer, florbetapir, was delayed until training methodologies were developed to reduce variability in interpreting amyloid- β plaque burden [7]. Although significant inter-reader variability remains (reported Kappa coefficients range 0.40–0.89), these FDA-approved methods have demonstrated adequate sensitivity for plaque detection with validation in cohorts of individuals with postmortem autopsies [6, 8–10]. Since the clinical implications of an elevated scan are significant, a highly-robust and reproducible method of interpretation is important. Development of meaningful quantitative PET imaging measures, such as using standard uptake value ratios (SUVRs), has an important clinical and research role, such as monitoring changes in amyloid- β plaque burden over time in longitudinal studies testing therapeutic interventions [11–12]. The development of a reliable binary interpretation measure (“elevated” vs. “non-elevated” amyloid- β plaque burden) remains the most clinically relevant measure for practicing radiologists and clinicians. Thus, the importance of an accurate and reproducible interpretation method of PET imaging is essential as the clinical use of amyloid- β PET scans becomes widespread.

Reliable detection of amyloid- β plaque prior to cognitive decline may be important, as new clinical treatment strategies that are currently in development that may disrupt or remove these plaques [13]. It is our experience that in patients with extensive plaque deposition, the visual interpretation of amyloid- β PET scans is usually not difficult. However, visual detection of plaque when confined to only a few areas is more challenging. Therefore, we developed and tested a systematic PET interpretation process that included both qualitative and quantitative assessment of amyloid- β plaque burden in an effort to improve the consistency of detection of early plaque deposition in cognitively normal older adults to identify individuals who may be at higher risk for developing AD.

MATERIALS & METHODS

We analyzed amyloid- β PET brain images from a small cohort of cognitively normal participants in ongoing AD prevention studies at our institution from August 1, 2011 to July 31, 2014. Subjects were administered intravenous florbetapir F-18 (370 MBq) and all

florbetapir PET images were obtained on a GE Discovery ST-16 PET/CT scanner. Two PET brain frames of five minutes in duration were acquired continuously, approximately 50 minutes post-injection of the florbetapir. Frames were then summed and attenuation corrected prior to interpretation.

Three experienced raters interpreted all images. Raters were a radiologist (rater 1), a neurologist (rater 2), and a neuroscience researcher (rater 3), each with at least one year of experience in florbetapir PET interpretation. Each rater completed the FDA-approved training program for florbetapir interpretation [14], which consisted of web-based lectures and training scans. MIMneuro software (MiM Software Inc, Cleveland, OH) was used to visualize, quantify plaque burden, and process the images.

Each rater participated in all three steps of the systematic PET interpretation process (Figure 1). First, the raters independently reviewed the raw amyloid- β PET images and qualitatively determined whether the scan was elevated or non-elevated (“Visual Read”). Visual Reads were completed according to the manufacturer’s training methods and prior to having access to or knowledge of quantitative analyses of amyloid- β plaque burden. Scans determined to be elevated on the Visual Read were defined as having either: 1) two or more areas of loss of grey-white matter differentiation larger than the cortical gyrus, or 2) one large area of grey matter uptake that clearly exceeded the adjacent white matter.

Second, MIMneuro software quantitatively compared the mean uptake in six regions of interest (ROIs) on the amyloid- β PET images to each patient’s mean whole cerebellum uptake, resulting in SUVRs for these six ROIs: anterior cingulate, posterior cingulate, precuneus, inferior medial frontal, lateral temporal, and superior parietal cortex. Each rater then used this quantitative data to classify each scan as elevated or non-elevated using a definition of two or more regions with mean SUVRs ≥ 1.1 (“Quantitative Read”). This cut point is consistent with previous imaging data using SUVRs of these same six regions that were correlated with autopsy measures [9–10]. Because the Visual Read required two or more areas of loss of grey-white matter differentiation, we chose a similar quantification cutoff of two or more of the six regions having SUVRs ≥ 1.1 .

In addition to providing SUVR analyses, MIMneuro also generated three supplementary analyses: region, voxel, and cortical projections (MIMneuro User Guide 6.2). In order to perform these analyses, each subject’s brain was mapped to a standard template space using a multi-template landmark-based deformable image registration (MIMneuro User Guide 6.2). After spatial normalization, each subject was compared on a voxel-by-voxel and regional basis to a database of 74 expert confirmed amyloid negative “normal” florbetapir-PET brain scans (48 males, 26 females) between the ages of 18–50. Z-scores, or the number of standard deviations away from the mean of the normal database, were calculated for each voxel in the brain and for the same six regions used in the SUVR analysis. Z-score results were displayed fused with the registered brain scan (Figure 1, Step 3, Panels A and B). Voxels with a z-score greater than three were color highlighted based on previous analyses demonstrating this value correlated most strongly with expert reads. Z-score values were also displayed for each of the six ROIs. Cortical surface projections were generated by sampling a probabilistic gray matter mask for the highest uptake voxel along vectors

perpendicular to the surface of the brain. Z-scores from the cortical projection images were also evaluated as an additional qualitative measure of potential affected cortical regions (MIMneuro User Guide 6.2).

Finally, after viewing all quantitative data, fused images, and cortical projections, each rater then reevaluated the raw PET images, paying close attention to ROIs demonstrating elevated SUVRs or z-scores. Raters documented their final interpretation of the amyloid- β PET scan as elevated or non-elevated in the context of both the visual and quantitative information ("VisQ Read"). Data were managed using REDCap electronic data capture tools hosted at our institution [15]. Statistical analyses were performed using SAS 9.4 software (SAS Institute, Inc., Cary, NC) and statistical significance was determined using a p-value <0.05. Cohen's Kappa was used to assess inter-rater agreement of reads and t-tests were used to compare continuous variables such as SUVR.

RESULTS

Three raters independently interpreted 55 amyloid- β PET scans from cognitively normal participants (one scan per participant) with an average age of 73.1 years ($SD \pm 5.9$), 56% female, and a mean mini-mental status exam (MMSE) score of 29 ($SD \pm 1.0$). The frequency of elevated Visual Reads ranged from 25.5% to 29.1% across raters (Table 1). Visual Read agreement across the three raters (Cohen's Kappa) ranged from 0.73 to 0.82 for this study. After re-interpreting the scans with quantitative information (VisQ Read), raters altered their initial interpretation in 16 (9.7%) of the 165 total amyloid- β PET reads, a frequency of change of 7.3% to 12.7% for individual rater's interpretations. Inter-rater agreement (Cohen's Kappa) for the final VisQ Reads improved from 0.88 to 0.96 after quantitative information was available. The majority of changes in interpretation went from non-elevated Visual Reads to elevated VisQ Reads (13 of 16 interpretations; 81.3%).

The 16 changed interpretations were from 11 participant scans. All three raters changed their interpretations for the same scan just once (from non-elevated to elevated; Figure 2 & 3), while two raters changed their interpretations for three scans and one rater changed interpretations for seven scans. The eight scans with non-elevated Visual Reads that changed to elevated VisQ Reads had a mean SUVR of 1.21 ($SD \pm 0.09$) and all had two or more ROIs with elevated SUVRs (Table 2). The scans with unchanged elevated reads ($n=12$) demonstrated a mean SUVR of 1.38 ($SD \pm 0.19$). None of the three scans with elevated Visual Reads that changed to non-elevated VisQ Reads had more than one ROI with an elevated SUVR, and their mean SUVR was 1.01 ($SD \pm 0.03$). This was similar to the 32 non-elevated scans for both the Visual and VisQ Reads (mean SUVR 0.99, $SD \pm 0.05$).

Based on quantitative information only (Quantitative Read), 45.5% ($n=25$) of the scans met criteria for elevated cerebral amyloid- β plaque (two or more ROIs with mean SUVRs ≥ 1.1). All non-elevated Quantitative Reads were non-elevated by both Visual and VisQ Reads (Figure 4). Of the 25 scans that were elevated by Quantitative Read, 44% ($n=11$) were not considered elevated by Visual Read and 24% ($n=6$) were not considered elevated by VisQ Read. This data suggests that Quantitative Reads alone may result in elevated interpretations more frequently than rater-dependent methods (Visual and VisQ Reads).

DISCUSSION

These data support our hypothesis that augmenting visual interpretation of amyloid- β PET scans of cognitively normal participants with quantitative information decreased the variability of rater interpretations, which is consistent with other recent studies utilizing a similar Visual Read only technique (Kappa 0.40–0.89) [6, 8–10, 12–13, 17–18]. Our results also suggest that the addition of quantitative PET scan data is most important for the improved detection of early plaque deposition or in cases where the patient demonstrates borderline Visual or Quantitative Reads (e.g., 1–2 qualitative and/or quantitative positive ROIs).

When quantitative information was provided to the raters, the frequency of an elevated cerebral amyloid- β interpretation rose slightly (from an average of 27.9% of Visual Reads to 36.4% of VisQ Reads) and inter-rater variability decreased, with VisQ Reads demonstrating higher inter-rater agreement (0.88 to 0.96), similar to a recent study utilizing a quantitative approach (0.92) [17]. Additionally, we did not find significant differences among the three raters with different clinical backgrounds (radiologist, neurologist, and neuroscience researcher), even though each had extensive experience with amyloid PET interpretation. This suggests that the three-step interpretation process may be robust enough for use across raters with different levels of imaging expertise.

Our study sample was composed of cognitively normal older adults with a prevalence of elevated cerebral amyloid- β plaque ranging from 25.5% at its lowest on Visual Reads to 45.5% at its highest on Quantitative Reads. This prevalence data aligns with estimates that approximately 30% of cognitively normal older adults have elevated cerebral amyloid- β plaque [19]. The broad variability in elevated amyloid- β plaque across these three different PET scan interpretation methods underscores the need for more precise interpretation processes. Although our study does not include long-term clinical outcomes or pathological confirmation of AD, the enhanced consistency across reads using the VisQ approach suggests this process provides a more stable measure than the Visual Read alone by reducing sources of variability inherent in subjective visual interpretation methods. This is particularly true for detection of early plaque deposition, as the PET scans with altered interpretations in this study had lower mean SUVRs than elevated scans that did not change.

The Quantitative Read estimate of 45.5% of individuals with elevated amyloid- β plaque is higher than expected in a cognitively normal study population. In at least five cases, elevated Quantitative Reads were unanimously determined by the raters to be non-elevated by both Visual and VisQ Reads. This suggests that the Quantitative Read threshold may be a good screening tool for detecting amyloid- β plaque, but not specific enough to make clinical inferences or management decisions. This is particularly important given the implications of an elevated or non-elevated PET scan for both individual patients and their clinicians. Further longitudinal clinical research is necessary to determine the relevance of elevated Quantitative Reads in the absence of visual elevations of amyloid- β plaque. Our cognitively normal sample population is unique for testing PET scan techniques for detecting early amyloid- β plaque deposition, as the prevalence of elevated plaque is lower than in

populations with mild cognitive impairment and dementia, although this could potentially limit the generalizability of these results to the general population.

This study suggests that interpretation of florbetapir PET scans should combine visual and quantitative analyses, similar to PET/CT oncologic imaging when standardized uptake values (SUVs) are used in combination with visual interpretation to make final diagnostic decisions. Because most clinical florbetapir PET scans will be interpreted by diagnostic radiologists with varying degrees of expertise in the field, a standardized process is necessitated. As in other medical fields, developing a systematic approach to the interpretation of these scans improves sensitivity and specificity, as well as reduces inter-rater variability. Our study findings suggest that interpretation approaches for florbetapir PET scans, particularly in the cognitively normal population, use a systematic process that combines both visual and quantitative information, which will uniquely minimize variability from a singular approach of using either a qualitative- or quantitative-only method.

Finally, the findings from this study may assist interpreting radiologists in three ways. First, if the Visual Read is initially non-elevated, careful analysis of the quantitative data is necessary and may change the final interpretation in approximately 10% of cases. This may be particularly important for inexperienced readers, who may utilize the quantitative data to direct their eye to subtle or equivocal regions of concern, in order to increase sensitivity and diminish variability between readers. Second, if the Visual Read is elevated and five or six regions are non-elevated on quantitative analysis, it is most likely a false positive visual read, as the non-elevated Quantitative Read alone demonstrates a high negative predictive value. Finally, quantitative-only reads appear to be overly sensitive and not specific enough to interpret alone without close visual inspection by a trained radiologist.

Specific limitations of this study are the relatively small sample size (n=55) and lack of a pathologic or tissue diagnosis as confirmation of the imaging findings. The relatively small sample will become more robust in anticipated longitudinal studies on the same individuals to evaluate for changes in their scan. There are obvious issues with obtaining a tissue biopsy on a cognitively normal participant, though, our results are at the expected frequency of elevated amyloid plaque reported elsewhere [19].

CONCLUSION

Our study suggests that combining both visual and quantitative information improves inter-rater variability of florbetapir PET scan interpretations in individuals without clinically meaningful cognitive change. This approach may lead to earlier detection of amyloid- β plaque, which may have implications for primary prevention interventions as therapeutics become more widely available.

Acknowledgments

Essential support and infrastructure was provided through P30 AG035982. Research reported in this publication was supported in part by R01 AG043962. Avid Radiopharmaceutical and Eli Lilly and Co. provided a grant to support the florbetapir imaging procedure.

References

1. Centers for Disease Control and Prevention. [Accessed 12/15/2015] Alzheimer's Disease. Mar. 2015 <http://www.cdc.gov/aging/aginginfo/alzheimers.htm>
2. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement*. 2011; 7:263–269. [PubMed: 21514250]
3. Mintun MA, Larossa GN, Sheline YI, et al. [11C]PIB in a nondemented population: Potential antecedent marker of Alzheimer's disease. *Neurology*. 2006; 67:446–452. [PubMed: 16894106]
4. Herholz K, Ebmeier K. Clinical amyloid imaging in Alzheimer's disease. *Lancet Neurol*. 2011; 10:667–670. [PubMed: 21683932]
5. Panegyres P, Berry R, Burchell J. Early dementia screening. *Diagnostics*. 2016; 6:6.
6. Lilly, Eli. [Accessed 12/15/2015] Amyvid United States Product Insert (USPI). Dec. 2013 <https://pi.lilly.com/us/amyvid-uspi.pdf>
7. Mintun, M. Image Interpretation and Training. Presented January 20, 2011 at FDA Advisory Committee Meeting; <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/PeripheralandCentralNervousSystemDrugsAdvisoryCommittee/UCM240951.pdf>
8. Center for Drug Evaluation and Research. [Accessed 12/15/2015] NDA 202008, Amyvid (Florbetapir, 18-F AV-45): Cross Discipline Team Leader Review. Mar. 2012 https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/202008Orig1s000CrossR.pdf
9. Clark CM, Pontecorvo MJ, Beach TG, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-plaques: A prospective cohort study. *Lancet Neurol*. 2012; 11:669–678. [PubMed: 22749065]
10. Joshi AD, Pontecorvo JM, Clark CM, et al. Performance characteristics of amyloid PET with florbetapir F-18 in patients with Alzheimer's disease and cognitively normal subjects. *J Nucl Med*. 2012; 53:378–384. [PubMed: 22331215]
11. Landau SM, Fero A, Baker SL, et al. Measurement of longitudinal β -amyloid change with 18F-florbetapir PET and standardized uptake value ratios. *J Nucl Med*. 2015; 56:567–574. [PubMed: 25745095]
12. Yamane T, Ikari Y, Nishio T, et al. Visual-statistical interpretation of 18F-FDG-PET images for characteristic Alzheimer patterns in a multicenter study: Inter-rater concordance and relationship to automated quantitative evaluation. *AJNR Am J Neuroradiol*. 2014; 35:244–249. [PubMed: 23907243]
13. Sevigny J, Chiao P, Bussière T, et al. The antibody aducanumab reduces A β plaques in Alzheimer's disease. *Nature*. 2016; 537:50–56. [PubMed: 27582220]
14. Lilly, Eli. [Accessed 12/15/2015] Amyvid Reader Training Program. Jan. 2014 <https://amyvid.myregistrationp.com/amyvid/login.do>
15. Harris P, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)--A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009; 42:377–381. [PubMed: 18929686]
16. Payoux P, Delrieu J, Gallini A, et al. Cognitive and functional patterns of nondemented subjects with equivocal visual amyloid PET findings. *Eur J Nucl Med Mol Imaging*. 2015; 42:1459–1468. [PubMed: 25952279]
17. Nayate AP, Dubroff JG, Schmitt JE, et al. Use of standardized uptake value ratios decreases interreader variability of [18F] florbetapir PET brain scan interpretation. *AJNR Am J Neuroradiol*. 2015; 36:1237–1244. [PubMed: 25767185]
18. Schreiber S, Landau SM, Fero A, et al. Comparison of visual and quantitative florbetapir F18 positron emission tomography analysis in predicting mild cognitive impairment outcomes. *JAMA Neurol*. 2015; 72:1183–1190. [PubMed: 26280102]
19. Chetelat G, La Joie R, Villain N, et al. Amyloid imaging in cognitively normal individuals, at-risk populations and preclinical Alzheimer's disease. *NeuroImage Clin*. 2013; 2:356–365. [PubMed: 24179789]

Figure 2. Example Florbetapir PET Scan Images

(A) Images from a “normal” scan with unchanged non-elevated Visual and VisQ Reads. (B) Images from a “borderline” scan in which reads changed from a non-elevated Visual Read to an elevated VisQ Read. (C) Images from an “abnormal” scan with unchanged elevated Visual and VisQ Reads. Left column depicts raw PET data, while right column depicts fused raw PET with Z-score quantitative data. This figure demonstrates the utility of quantitative analysis for detecting subtle abnormalities in the Visual Read and raw PET images, particularly for patient B, in whom the interpretation was changed by all three raters.

Figure 3. Histogram of Cumulative VisQ and Visual Reads

The VisQ Read is the final interpretation with agreement between at least two raters, either elevated or non-elevated. The Visual Read represents where consensus was reached by at least two of the three raters as either elevated or non-elevated. The mean SUVR of the six analyzed cortical regions for each participant is listed in the respective subset, with 1.10 used as the cutoff between an elevated and non-elevated region.

Table 1

Frequency of Elevated Reads and Inter-Rater Variability

Rater	Visual Read		VisQ Read		Read Changes
	Elevated% (n)	Consistency	Elevated% (n)	Consistency	
1	29.1% (16)	0.82 *	34.6% (19)	0.96 *	9.1% (5)
2	25.5% (14)	0.73 **	32.7% (18)	0.88 **	7.3% (4)
3	29.1% (16)	0.74 ***	34.6% (19)	0.92 ***	12.7% (7)

Consistency indicates Cohen's Kappa Value.

*
rater 1 vs 2,

**
rater 2 vs 3,

rater 3 vs 1

Table 2

Visual & VisQ Reads Compared to Quantitative Read Mean SUVR

	Elevated Visual & VisQ Read	Non-Elevated Visual & VisQ Read	Visual Read Elevated, VisQ Read Non-Elevated	Visual Read Non-Elevated, VisQ Read Elevated
Number of Participants (n)	12	32	3	8
Number of elevated regions (SD)	5.5 (1.0)	0.6 (1.0)	0.7 (0.6)	4.6 (1.5)
Mean SUVR of 6 ROIs (SD)	1.38 (0.19)	0.99 (0.05)	1.01 (0.03)	1.21 (0.09)