

Below we provide a draft description of results of one of the research questions as set in the Flutemetamol study (trial registry link <https://www.clinicaltrialsregister.eu/ctr-search/trial/2012-002303-18/NL>). During the project the validity of the method was discussed, in light of the swiftly developing state of the art in the field. After consultation of an external expert the decision was made not to finalize the manuscript and refrain from submitting the work to a journal.*

** Prof. Dr. Frisoni from Memory Clinic, Geneva University and University Hospitals, Geneva, Switzerland*

Added diagnostic accuracy after [18F]flutemetamol PET scanning in young patients with dementia in an academic memory clinic setting

CONTRIBUTORS

Ron L.H. Handels¹

Marissa D. Zwan²

Wiesje Pelkmans²

Geert Jan Biessels³

Peter van Domburg⁴

Erik Hoff⁵

Gerwin Roks⁶

Niki Schoonenboom⁷

Niek Verweij⁸

Bart N.M. van Berckel²

Wiesje M. van der Flier²

Frans Verhey¹

Philip Scheltens²

Femke H. Bouwman²

AFFILIATIONS

¹ Alzheimer Centre Limburg, School for Mental Health and Neuroscience, Department of Psychiatry and Neuropsychology, Maastricht University, 6200 MD, Maastricht, The Netherlands.

² Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, VU University Medical Center, Amsterdam UMC, The Netherlands

³ University Medical Centre Utrecht, 3508 GA Utrecht, The Netherlands.

⁴ Zuyderland, location Sittard-Geleen, 6130 MB Sittard-Geleen, The Netherlands.

⁵ Zuyderland, location Heerlen, 6130 MB Sittard-Geleen, The Netherlands.

⁶ Elisabeth-TweeSteden Ziekenhuis, 5000 LC Tilburg, The Netherlands.

⁷ Spaarne Gasthuis, 2000 AK Haarlem, The Netherlands.

⁸ Medisch Centrum Leeuwarden, 8901 BR, The Netherlands.

ABSTRACT

Introduction: A timely diagnosis of Alzheimer's Disease (AD) in an early stage of dementia is important to support timely access to treatment, advice, and care. The aim of this study was to estimate the diagnostic accuracy of [18F]flutemetamol PET in addition to the usual diagnostic workup for the diagnosis of AD in a memory clinic population with young onset dementia by means of a panel reference-based etiology diagnosis.

Methods: in an academic memory clinic early onset dementia cohort (n=211) the nosological diagnosis was set by usual diagnostic workup and after including [18F]flutemetamol amyloid PET in a stepwise approach. To assess the change in proportion correctly diagnosed, the diagnosis with and without [18F]flutemetamol PET was related to a panel-based reference standard, serving as gold standard, consisting of 3 neurologists who relied on available clinical information over 2-year follow-up (n=152; blinded for PET).

Results: The panel majority nosology was set as a reference diagnosis in 122 participants, leaving 30 (20%) participants with no majority reached. In 107 (88%) cases post-PET was in line with the reference, and in 103 (84%) the pre-PET diagnosis was in line with the reference. The difference was 3.3% (95% CI -3.5% to 10.1%; p=0.424).

Discussion: [18F]flutemetamol PET did not significantly improve the diagnostic accuracy in young patients with dementia in an academic memory clinic setting. The secondary analyses provided several indications for future research in a narrower subsample of persons with (very) high diagnostic uncertainty and to assess patient relevant health outcomes.

INTRODUCTION

A timely diagnosis of Alzheimer's Disease (AD) in an early stage of dementia is considered important to support timely access to treatment, advice, and care [ADI report 2011; Dubois, 2016].

In the last decades diagnostic research has focused on the use of amyloid biomarkers in CSF and on PET to identify amyloid pathology in dementia as well as pre-dementia stages of mild or subjective cognitive impairment, which has led to the development of new clinical and research criteria [Jack, 2018; Albert, 2011; McKhann, 2011].

Fluorine-18 labeled amyloid tracers, among which [18F]flutemetamol with a relatively long half-life, enabled a wide spread application of this amyloid biomarker to serve clinics outside its tracer synthesis facility as compared to for example the [11C] Pittsburg Compound-B (PIB) tracer.

Johnson et al. [2016] indicated the clinical use of amyloid PET is appropriate when the possible diagnosis of AD is uncertain after a comprehensive evaluation by a dementia expert and knowledge of the presence or absence of Amyloid pathology is expected to increase diagnostic certainty and alter management. In addition, recommendations to use and reimburse amyloid PET should rely on its clinical utility in terms of empirical evidence on health outcomes and evidence on cost-effectiveness of the test-treat or test-care pathway [Lijmer, 2009a].

In terms of health outcomes, systematic reviews on the effect of early identification of AD hallmarks did not find empirical evidence of health benefits derived from persons with and without early identified AD genetic profile or pathological hallmarks, but nevertheless reported on potential benefits and harms of early AD testing across the clinical stages of pre-clinical, MCI and dementia in the absence of (disease-modifying) treatment [Dubois, 2016; Paulsen, 2013; ADI, 2011].

In terms of diagnostic accuracy, ideally, a gold standard is available to determine the target condition. However, as this is not the case for AD [Scheltens, 2011] a next best alternative is to rely on a clinical reference standard. A variety of reference standards have been used to compare the diagnostic performance of the clinical practice diagnostic workup without and with using the results from PET imaging. Some studies evaluated amyloid PET in terms of progression from MCI to clinically established AD-type dementia after 2 years, using clinical assessment [Ossenkuppele, 2012] or statistical modelling [Vos, 2015]. Some studies used neuropathology at autopsy [Curtis, 2015; Sanchez-Juan, 2014]. Various studies have not used a reference standard and focused on the proportion of change in clinical diagnosis, mean change in diagnostic certainty and/or proportion clinical (treatment) management before and after disclosing the PET result using various [18F] amyloid tracers (florbetapir, flutemetamol, florbetaben) [Ceccaldi, 2018; Leuzy, 2018; Zwan, 2017; Rabinovici, 2019; Trivino, 2019]. Although this evidence is highly valuable and supports diagnostic test development, a reference standard consisting of conversion from MCI to dementia is mainly limited to reflecting the prognostic value of PET. A reference standard consisting of pathology is limited by disregarding the clinical expression of the underlying cause. A focus on diagnostic change rather than comparing to a reference standard is limited by relying on the assumption that any change is a correct change.

In absence of a gold standard, the use of a panel has been recommended in order to combine multiple test results and construct a reference standard outcome [Rutjes, 2007].

To the best of our knowledge only Hellwig et al. [2018] assessed Amyloid PET in terms of its relation to a clinically relevant reference standard. They used [11C]PIB amyloid PET in 84 patients with a major neurocognitive disorder of uncertain etiology in an academic memory clinic. An etiology diagnosis was set before and after interpreting PET in addition to relevant clinical information (neuropsychological assessment, MRI/CT, CSF). The reference standard consisted of an interdisciplinary board consensus on etiology relying on all available clinical information on clinical symptoms, neuropsychological testing, additional imaging and CSF (from pre-PET and mean 2.4 year follow-up), as well as baseline [11C]PIB and [18F]FDG PET information.

With no longitudinal evidence available on [18F]flutemetamol PET the aim of this study was to estimate the diagnostic accuracy of [18F]flutemetamol PET in addition to the usual practice diagnostic workup for the clinical diagnosis in a memory clinic population of patients with young onset dementia by means of a panel reference-based etiology diagnosis that uses clinical follow-up information. We hypothesize a larger proportion of correctly diagnosed participants after PET compared to the diagnosis before PET, according to the reference standard.

METHODS

PARTICIPANTS

In the Dutch Flutemetamol Study, participants received an [18F]flutemetamol PET scan and were followed up for 2 years in a longitudinal prospective cohort design. Participants were recruited between 2012 and 2015 from the VU University Medical Center (n=200) and Maastricht University Medical Centre+ (n=11) if they were suspected of dementia, aged ≤ 70 and mentally competent (MMSE ≥ 18) and diagnostic certainty between 50-90% (see supplemental 1 and Zwan et al. [2017] for details). Written informed consent was obtained from all participants. This study was approved by the medical ethics review committee of the VU University Medical Center (reference number 2012/302).

Participants were omitted from the analysis due to insufficient follow-up information on cognition, behavior and activities of daily living (ADL) (either obtained from test results or from (telephone) interview), or on imaging or biomarker information (see figure 1 for reasons of drop-out).

CLINICAL ASSESSMENTS

At baseline, participants received a standardized clinical dementia evaluation including information on demographics, medical history, informant-based history, physical examination, neurological examination, neuropsychological examinations, screening laboratory tests, brain MRI, and neuropsychological testing [van der Flier, 2014; Aalten, 2014]. In the absence of contraindications, a lumbar puncture was performed and additional investigations considered relevant (e.g. FDG-PET, psychiatric assessment, blood analysis). A standardized clinical follow-up was performed after 1 and 2 years, at which the same information was obtained except from imaging, CSF and blood samples.

If a person was not able to attend the memory clinic for an assessment, a telephone interview was performed to obtain as much information as possible on the presence of dementia-related symptoms and progression.

See supplemental 2 for an example vignette with hypothetical diagnostic information.

[18F]flutemetamol PET

In both centers, [18F]flutemetamol PET scans were made on a Gemini TF-64 PET/CT scanner (Philips Medical Systems, Best, the Netherlands) [Surti, 2007]. Ninety minutes after a bolus injection of 191 ± 10 MBq [18F]flutemetamol, patients underwent a low-dose CT scan followed by a 20-minute (i.e., 4 frames of 5 minutes) PET scan. Scans were checked for movement and frames were summed to obtain a static (20-minute) image for each patient (except for one patient in whom the last frame was not used due to extensive head movement). Scans were visually assessed and dichotomously rated as either amyloid-positive or amyloid-negative by the local nuclear medicine physician, who completed the training program for visual interpretation of [18F]flutemetamol images. Readers were blinded to clinical information, except for brain MRI.

PRE- AND POST-PET ETIOLOGY DIAGNOSIS

At baseline, an etiology diagnosis was set in a multidisciplinary meeting according to established criteria [McKhann, 2011; Rascovsky, 2011; Roman, 1993; McKeith, 2005; Boeve, 2003; Litvan, 1996] based on the diagnostic information obtained in usual practice, except for altering the usual practice by blinding them for CSF results (further referred to as pre-PET etiology diagnosis).

After about 4-12 weeks a [18F]flutemetamol PET scan was performed and rated as AD or non-AD by a specialist in nuclear medicine. The treating neurologist combined this information with the pre-PET diagnostic information to set an etiology diagnosis, guided by AD diagnostic research criteria [McKhann, 2011] (further referred to as post-PET etiology diagnosis).

For the purpose of this study, lumbar puncture results were not disclosed before the impact of PET results had been assessed. All etiology diagnoses were rated as AD, vascular dementia (VAD), frontotemporal dementia (FTD), Dementia with Lewy bodies (DLB), other neurodegenerative disease (other NDD) and no neurodegenerative disease (no NDD) as well as their certainty, using standardized forms (see supplemental 3).

REFERENCE ETIOLOGICAL DIAGNOSIS

A similar method to Handels et al. [2016] was followed and recommendations for designing a panel diagnosis [Bertens, 2013; Handels, 2014] were taken into account as much as possible.

All information described in the paragraph ‘clinical assessments’ was summarized on a digital vignette, which was rated online independently by 3 neurologists (among GB, PD, EH, GR, NS and NV) with over [x] years of clinical experience operating in a local (PD, EH, GR, NS and NV) or university hospital (GB). Ratings were done blinded for [18F]flutemetamol PET ensuring no diagnostic review bias. They individually rated for each etiology (AD, VAD, FTD, DLB, otherNDD, noNDD) if it determined the clinical picture on

a 15-point scale (see supplemental 3). Each panelist's set of ratings was standardized by dividing it by his/her total attributed points to all etiologies. If at least 2 panelists had identified the same etiology or identical combination of etiologies with a standardized rating of ≥ 0.33 (i.e. one third of the total), this was set as the majority etiology (see supplemental 4 for examples). If not, it was set as inconclusive. Alternatively, the panelists' ratings of the inconclusive cases were interpreted by FB who, blinded for PET, identified the most plausible etiology if possible. This was meant to identify plausible evident etiologies that could not be identified by the majority algorithm. If still not possible, it was considered inconclusive.

[Comment from Prof. Dr. Frisoni: "I am not sure I agree that a consensus diagnosis based on clinical data and a 2-year follow-up is an appropriate reference. In patients with MCI, follow-up can elucidate converters from non-converters (although 2 years is a short span). In patients with dementia, the phenotype is not going to change in 2 years (and not even in 5 for that matter). AD phenotypes will stay AD and FTD phenotypes will stay FTD. Those AD who develop BPSD do not necessarily have FTD, and FTD who develop memory problems do not necessarily have AD. The notion of using a clinical phenotype as the reference dx goes against the modern notion of etiological diagnosis based on the molecular profile. I suggest to use the consensus dx only as a validation of the pre-PET dx by the local neurologist, not of the post-PET dx. This approach would IMHO save the dataset and the enormous amount of effort that went into the dx consensus (instead this is a strength of the dataset – if well exploited)."]

STATISTICAL ANALYSIS

Demographic characteristics and the changes in etiology diagnosis from pre-PET to post-PET stratified by the reference etiology diagnosis were described.

For descriptive purposes the etiological diagnosis was classified as AD (including mixed AD) or non-AD. PET was considered to have had an added diagnostic value if the post-PET etiology diagnosis differed from the pre-PET etiology diagnosis and the post-PET etiology was in concordance with the reference etiology diagnosis. PET was considered to have generated a loss in diagnostic value if the post-PET etiology diagnosis differed from the pre-PET etiology diagnosis, and the post-PET etiology diagnosis was different from the reference etiology diagnosis. In all other situations, PET was considered to have no added diagnostic value and did not result in a loss in diagnostic value. Inconclusive reference outcomes were omitted from the analyses and described separately.

The proportion correct etiology diagnoses were tested both for pre-PET and post-PET using a McNemar's chi-squared test using STATA15 (significance level set at 0.05). The power of the analysis was calculated using a power calculation for paired proportions, and minimal detectable difference under 80% power.

SECONDARY AND SENSITIVITY ANALYSES

Several secondary and sensitivity analyses were performed to explore the data and assess the robustness of the results.

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) was estimated for both the pre- and post-PET dichotomized etiology diagnosis in terms of AD versus non-AD.

The level of importance panelists rated the different categories of diagnostic information was described.

The association between baseline characteristics and correctness of the diagnostic change was estimated (i.e. to explore predicting in which individuals PET likely has an added value) using bivariate logistic regression analyses.

In a sensitivity analysis the McNemar's test was performed using the certainty of the reference diagnosis as quasi weights.

The ratings from 3 panelists on etiology for those participants in which no majority etiology was reached and no plausible evident etiology could be set were listed.

The primary and part of the secondary analyses were performed both in- and excluding the plausible evident reference etiology diagnoses.

Due to the explorative nature of the sensitivity analyses, the significance levels were not adjusted for multiple testing.

RESULTS

Part of the 1-year follow-up (n=4) and the 2-year follow-up (n=46) assessments were not performed leaving 102 participants with 2 follow-up and 50 participants with 1 follow-up assessment for analysis (see figure 1). Age, sex and education were similar, but MMSE (mean 23.8 versus 22.4), DAD (mean 85 versus 76) and NPI (mean 12 versus 19) were significantly more affected in the omitted part of the sample.

The average age was 62 years (range: 45-70) and 44% was female. The mean MMSE was 23.8 (SD=3.5). See table 1 for details.

A reference diagnosis was set in 122 participants (80%) with 56 a full and 66 a partly majority diagnosis (of whom 22 and 18 only had one of out of the 2 follow-up assessments available respectively). In 30 (20%) participants no majority was reached. Of these 30, in 22 a plausible evident etiology was set, leaving 8 participants with a definite non-majority etiology.

Of the subsample with full or partly majority diagnosis, in 107 (88%) cases post-PET was in line with the reference, and in 103 (84%) the pre-PET diagnosis was in line with the reference. The difference in proportions was 0.033 (95% CI -0.035 to 0.101; p=0.424). The power to detect this difference with the available sample size was 0.22 (a difference of 0.075 could be detected with the available sample size). Using the plausible evident etiology 122 post-PET (85%) and 118 pre-PET (82%) were in line with the reference (difference 0.028; 95% CI -0.037 to 0.092; p= 0.481; power=0.16; detectable difference with sample size was 0.077).

In 20 cases (16%) the diagnosis changed between pre-PET to post-PET. Of those 20, 9 changes were correct (i.e. post-PET in line with reference), 7 changes were incorrect (i.e. post-PET differed from reference), 4 changes were within a mixed reference etiology (e.g. pre-PET AD to post-PET FTD, majority etiology was mixed AD/FTD). When including the

plausible evident cases, 27 (19%) changes occurred of which 11 correct, 11 incorrect and 5 within mixed.

See table 2 for the details stratified by the reference diagnosis.

Figure 2 displays the pre- and post-PET diagnostic certainty classified by correct, incorrect or no change in etiological diagnosis. In a large proportion of the participants the diagnostic certainty increased.

SECONDARY AND SENSITIVITY ANALYSES

Pre-PET sensitivity and specificity were 0.89 and 0.70 and post-PET were 0.85 and 0.92 respectively. Sensitivity did not significantly change, while specificity did significantly change ($p=0.031$). This was 0.86 and 0.69 pre-PET, and 0.83 and 0.92 post-PET respectively when including plausible evident etiology ratings, also with significant specificity change.

PPV and NPV was 0.91 and 0.66 for pre-PET and 0.98 and 0.64 for post-PET respectively. This was 0.89 and 0.63 for pre-PET, and 0.97 and 0.65 for post-PET when including plausible evident etiology ratings.

The mean (standard deviation) level of importance ratings from the panelists of the different categories of diagnostic information on a scale from 1 to 15 was 11.1 (2.5) for clinical information, 9.9 (4.7) for CSF, 10.1 (3.0) for MRI, 9.8 (2.7) for neuropsychiatric examination and 5.4 (5.7) for other information (such as blood test or FDG PET).

Bivariate logistic regression on correct versus incorrect or no change (with change within mixed classified as no change) using age, gender, education, MMSE, CDR, GDS, DAD, dichotomized pre-PET diagnosis (AD/non-AD) and pre-PET diagnostic certainty indicated only pre-PET diagnosis not being AD was significantly associated to a correct change after PET (both without and with the plausible evident cases). An alternative analysis classifying change within mixed as change indicated lower CDR and lower pre-PET certainty were significantly associated to this (and in addition to these 2 also pre-PET diagnosis not being AD when using plausible evident cases).

The quasi-weighted McNemar's test indicated the proportion correctly diagnoses was not significantly different.

Supplemental 5 describes the ratings from 3 panelists on etiology for those participants in which the no majority etiology diagnosis was reached and no plausible evident etiology diagnosis could be set.

DISCUSSION

In a cohort of 122 young onset dementia patients in an academic memory clinic [18F]flutemetamol PET performed after the usual diagnostic workup resulted in a correct change of the etiological diagnosis in 9 participants and an incorrect change in 7 participants, as compared to a panel reference etiology using up to 2 year follow-up information on clinical symptoms. The percentage correctly diagnosed was 84% pre-PET and 88% post-PET (not significant).

In this study [18F]flutemetamol PET did not result in a significant improvement of correct diagnostic changes. However, the statistical power was insufficient to establish the significance of the relatively small difference that was observed (4% improvement). An

explanation why PET did not result in a larger improvement is that the standard diagnostic workup already includes an extensive set of diagnostic information on patient history, and physical, neurological, psychiatric, neuropsychological and MRI. CSF was obtained in part of the sample (xx%), but the rated the pre- and post-PET diagnosis were rated blinded for CSF results. The added value of PET was possibly further limited if CSF would have been part of the pre-PET diagnostic workup, as amyloid PET likely is associated to CSF amyloid beta.

In addition, mean pre-PET certainty on the etiology was 70 (on a scale of 1-100; SD=12, and the highest quartile ranged from 79 to 89), which could be considered relatively high. Post-hoc analyses on a subsample of those with 50 to 65 pre-PET uncertainty indicated the percentage correctly diagnosed was 71% in pre-PET and 79% in post-PET. This indicates a possible higher value in more uncertain cases. The secondary analysis (including the plausible evident cases) further support this as it indicated that pre-PET diagnosis not being AD, a lower pre-PET certainty as well as less affected cognition and function (measured by the CDR) were associated to improved correct diagnostic changes. The less affected symptoms might indicate a possible added value in persons with very mild dementia although this was not confirmed by the secondary analysis on function measured by the DAD and cognition measured by the MMSE. The possible higher added value in more uncertain cases is further supported by the findings of Hellwig et al. [2018] who assessed a sample of patients with a complicated clinical presentation. They found a lower pre-PET accuracy of 71% (compared to 84% in our study) and a larger proportion of 23% in whom the diagnosis was changed (compared to 16% in our study), arriving at a post-PET accuracy of 89% (compared to 88% in our study). The plausibility that PET has a larger added value in uncertain cases is further supported by the study by Ceccaldi et al. [2018] who reported a large change in diagnosis (67%) in a sample of patients in whom CSF was not possible or even after CSF testing the etiology remained uncertain. However, opposite to the study by Hellwig et al., this study did not assess the accuracy by means of a (panel) reference standard. Figure 2 is also in line with this explanation as it shows only 1 diagnostic change in those with a pre-PET certainty of 80 or higher, reflecting no added diagnostic accuracy value (other than the potential value of increased certainty).

The vignettes used in the reference standard were composed of both clinical symptoms (phenotypical expression) as well as pathological measures (i.e. MRI and CSF), although longitudinal information on pathological measures was only available in a few cases. Therefore, the [18F]flutemetamol PET scan result could have correctly identified amyloid pathology while being discordant with a non-AD reference outcome. Postmortem studies have shown AD and DLB or AD pathology can be both present [Harding, 2001], and clinical expressions of FTD with AD pathology present [Johnson, 1999]. Although this seems contradicting, it is in line with the purpose of this study to assess the clinical utility of PET (rather than its correlation to pathology) as clinical management is oriented towards current but also future symptoms, such as practical support in home setting, treatment and case management.

Obviously in case of future anti-amyloid therapy it becomes essential to assess PET in terms of its association to pathology, which has indicated a relatively high accuracy [Curtis, 2015].

Alternatively, the [18F]flutemetamol PET result could have been included in the vignettes for the reference standard, as has been done by Hellwig et al. [2018]. This could have improved the reference standard, especially for atypical representations. It would likely have resulted in a different reference diagnosis in the 8 cases with ‘no change (incorrect)’ (see table 2). However, its importance could be overestimated causing incorporation bias and leading to an

overestimation if the PET's accuracy [whiting, 2004] and therefore often advised not to [Rutjes, 2007].

STRENGTHS AND LIMITATIONS

This study relied on a reference standard based on longitudinal data, which generated empirical evidence in a relatively large sample of 122 participants on the correctness of the diagnostic changes due to [18F]flutemetamol PET in an academic setting.

Our study was limited due to several reasons. Drop-out was relatively large (28%) and diagnostic changes occurred less in the sample remaining for analysis (16% instead of 19% in the total sample [Zwan, 2017]). This could have resulted in a drop-out bias underestimating the clinical utility of [18F]flutemetamol PET. However, of the 3 factors associated with a correct diagnostic change (CDR, pre-PET diagnostic certainty and pre-PET diagnosis of AD) only CDR was higher in the omitted sample (0.84 versus 0.71 in the sample remaining for analysis), which makes a large underestimation unlikely.

We have altered usual practice by excluding CSF from the pre- and post-PET diagnosis. It is expected that if amyloid beta CSF was included in the pre-PET diagnosis, it would have reduced the added diagnostic value of PET due these both amyloid tests being correlated. Nevertheless, the current estimates reflect the effect of PET if a lumbar puncture could not be performed due to for example contra-indications, technical problems or patient refusal.

The follow-up period of 2 years possibly limited a confident reference diagnosis in participants with slow disease progression as for example AD pathology could come to expression after 2 years (i.e. imperfect reference standard bias). However, a too long follow-up up to a decade would drive the reference standard towards a pure prognostic marker as it becomes less likely that longer-term symptoms retrospectively apply to the baseline situation (i.e. disease progression bias: when index test is performed unusually long before the reference standard, such that the disease is more advanced when the reference standard is performed).

A panel reference standard requires a large time investment of panelists, and would be even larger if a stepwise evaluation was incorporated (e.g. to assess the value of including PET in the reference standard by subsequently providing [18F]flutemetamol PET results in the patient vignettes). Alternative to a panel reference standard a composite or latent reference standard could be determined based on all information from the vignette [Bertens, 2013]. However, this would omit the information from clinical history and is complex as the vignette includes a large variety of detailed information. Due to time constraints some recommendations were omitted [Bertens, 2013] possibly limiting the validity of the study: including 3 panelists instead of 5 or 7, no validation assessment in a subsample, no video recordings to see the participant history interview 'live', and no (staged) discussion meeting to reach consensus on non-majority cases [Bertens, 2014]. As an alternative to the latter, the panelists ratings of each of the 8 non-majority cases were assessed by a single clinician (FB, blinded for PET and not being one of the panelists). As the analyses including these data showed similar results to the analyses excluding this approach showed similar results, the conclusion is likely unaffected by this design choice. In addition, we believe that a forced consensus on cases with large disagreement among panelists likely results in arbitrary reference diagnoses. However, these are probably important cases as the proportion of change was larger in the 8 cases without a reference etiology diagnosis available (50%) than in the 122 cases with a reference etiology diagnosis available (16%).

CLINICAL AND RESEARCH IMPLICATIONS

Based on the primary outcome of no increased significant diagnostic accuracy our study does not support the implementation of [18F]flutemetamol amyloid PET in routine clinical academic practice.

However, based on the secondary and post-hoc analyses we hypothesize a benefit of amyloid PET (in terms of increased diagnostic accuracy when compared to a reference standard) in academic clinical practice in persons with low pre-PET diagnostic certainty.

Diagnostic accuracy remains an intermediate outcome for the final goal of improving patients' health or limit health loss. Therefore, future studies should focus beyond this concept and assess health outcomes even in absence of disease-modifying treatments [Bossuyt, 2009], preferably in a randomized setting with and without (disclosing) amyloid PET [Lijmer, 2009b].

CONCLUSION

[18F]flutemetamol amyloid PET did not significantly improve the diagnostic accuracy in young patients with dementia in an academic memory clinic. The secondary analyses indicated the potential added value of amyloid PET in a narrower subsample of persons with (very) high diagnostic uncertainty, which we recommend to assess in future research.

ACKNOWLEDGMENT

This study was performed within the framework of the Dutch Flutemetamol Study and supported by the Dutch Alzheimer's Society (grant WE.15-2014-01) and through an unrestricted grant of GE Healthcare to the Stichting Alzheimer & Neuropsychiatrie, Amsterdam. Research of the VUmc Alzheimer Center is part of the neurodegeneration research program of the Neuroscience Campus Amsterdam. The VUmc Alzheimer Center is supported by Alzheimer Nederland and Stichting VUmc fonds. The clinical database structure was developed with funding from Stichting Dioraphte.

Dutch Flutemetamol Study EudraCT Number 2012-002303-18 (clinicaltrialsregister.eu).

Prof. Dr. Frisoni has provided written consent to publish his comment and name as provided in this manuscript.

REFERENCES (draft)

- Aalten P, Ramakers IH, Biessels GJ, de Deyn PP, Koek HL, OldeRikkert MG, Oleksik AM, Richard E, Smits LL, van Swieten JC, Teune LK, van der Lugt A, Barkhof F, Teunissen CE, Rozendaal N, Verhey FR, van der Flier WM. The Dutch PARELSNOER Institute-- Neurodegenerative diseases; methods, design and baseline results. *BMC Neurol*. 2014 Dec 31;14:254. doi: 10.1186/s12883-014-0254-4. PMID: 25551191; PMCID: PMC4301568.
- ADI 2011. Prince M, Bryce R, Ferri CP (2011) World Alzheimer Report 2011. Alzheimer Disease International, London.
- Bertens LC, Broekhuizen BD, Naaktgeboren CA, Rutten FH, Hoes AW, van Mourik Y, Moons KG, Reitsma JB. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS Med*. 2013 Oct;10(10):e1001531. doi: 10.1371/journal.pmed.1001531. Epub 2013 Oct 15. PMID: 24143138; PMCID: PMC3797139.
- Bertens LC, van Mourik Y, Rutten FH, Cramer MJ, Lammers JW, Hoes AW, Reitsma JB, Moons KG. Staged decision making was an attractive alternative to a plenary approach in panel diagnosis as reference standard. *J Clin Epidemiol*. 2015 Apr;68(4):418-25. doi: 10.1016/j.jclinepi.2014.09.020. Epub 2014 Nov 4. PMID: 25459980.
- Boeve BF, Lang AE, Litvan I. Corticobasal degeneration and its relationship to progressive supranuclear palsy and frontotemporal dementia. *Ann Neurol*. 2003;54:S15-9.
- Bossuyt PM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. *Med Decis Making*. 2009 Sep-Oct;29(5):E30-8. doi: 10.1177/0272989X09347013. Epub 2009 Sep 2. PMID: 19726782.
- Ceccaldi M, Jonveaux T, Verger A, Krolak-Salmon P, Houzard C, Godefroy O, Shields T, Perrotin A, Gismondi R, Bullich S, Jovalekic A, Raffa N, Pasquier F, Semah F, Dubois B, Habert MO, Wallon D, Chastan M, Payoux P; NEUUS in AD study group, Stephens A, Guedj E. Added value of 18F-florbetaben amyloid PET in the diagnostic workup of most complex patients with dementia in France: A naturalistic study. *Alzheimers Dement*. 2018 Mar;14(3):293-305. doi: 10.1016/j.jalz.2017.09.009. Epub 2017 Nov 4. PMID: 29107051.
- Curtis C, Gamez JE, Singh U, et al. Phase 3 trial of flutemetamol labeled with radioactive fluorine 18 imaging and neuritic plaque density. *JAMA Neurol* 2015;72:287-94.
- Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G (2015) Timely diagnosis for Alzheimer's disease: A literature review on benefits and challenges. *J Alzheimers Dis* 49, 617-631.
- van der Flier WM, Pijnenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, van Berckel BN, Stam CJ, Barkhof F, Visser PJ, van Egmond E, Scheltens P. Optimizing patient care and research: the Amsterdam Dementia Cohort. *J Alzheimers Dis*. 2014;41(1):313-27. doi: 10.3233/JAD-132306.
- Gabel MJ, Foster NL, Heidebrink JL, Higdon R, Aizenstein HJ, Arnold SE, Barbas NR, Boeve BF, Burke JR, Clark CM, Dekosky ST, Farlow MR, Jagust WJ, Kawas CH, Koeppe RA, Leverenz JB, Lipton AM, Peskind ER, Turner RS, Womack KB, Zamrini EY. Validation of consensus panel diagnosis in dementia. *Arch Neurol*. 2010 Dec;67(12):1506-12. doi: 10.1001/archneurol.2010.301. PMID: 21149812; PMCID: PMC3178413.
- Handels RL, Wolfs CA, Aalten P, Bossuyt PM, Joore MA, Leentjens AF, Severens JL, Verhey FR. Optimizing the use of expert panel reference diagnoses in diagnostic studies of

multidimensional syndromes. *BMC Neurol.* 2014 Oct 4;14:190. doi: 10.1186/s12883-014-0190-3. PMID: 25280531; PMCID: PMC4195860.

Handels RL, Joore MA, Vos SJ, Aalten P, Ramakers IH, Rikkert MO, Scheltens P, Jansen WJ, Visser PJ, van Berckel BM, van Domburg P, Smid M, Hoff E, Hoogmoed J, Bouwman F, Claassen J, Leentjens AF, Wolfs CA, Severens JL, Verhey FR. Added Prognostic Value of Cerebrospinal Fluid Biomarkers in Predicting Decline in Memory Clinic Patients in a Prospective Cohort. *J Alzheimers Dis.* 2016 Mar 31;52(3):875-85. doi: 10.3233/JAD-151120. PMID: 27031483.

Harding AJ, Halliday GM. Cortical Lewy body pathology in the diagnosis of dementia. *Acta Neuropathol* 2001;102:355–63.

Hellwig S, Frings L, Bormann T, Vach W, Buchert R, Meyer PT. Amyloid imaging for differential diagnosis of dementia: incremental value compared to clinical diagnosis and [18F]FDG PET. *Eur J Nucl Med Mol Imaging.* 2019 Feb;46(2):312-323. doi: 10.1007/s00259-018-4111-3. Epub 2018 Aug 10. PMID: 30094462; PMCID: PMC6333717.

Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R; Contributors. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* 2018 Apr;14(4):535-562. doi: 10.1016/j.jalz.2018.02.018.

Jiménez-Bonilla JF, Quirce R, De Arcocha-Torres M, Martínez-Rodríguez I, Martínez-Amador N, Sánchez-Salmón A, De la Fuente-Gómez F, Rodríguez E, Sánchez-Juan P, Banzo I. A 5-year longitudinal evaluation in patients with mild cognitive impairment by 11C-PIB PET/CT: a visual analysis. *Nucl Med Commun.* 2019 May;40(5):525-531. doi: 10.1097/MNM.0000000000001004. PMID: 30973840.

Johnson KA, Minoshima S, Bohnen NI, Donohoe KJ, Foster NL, Herscovitch P, Karlawish JH, Rowe CC, Carrillo MC, Hartley DM, Hedrick S, Pappas V, Thies WH; Alzheimer's Association; Society of Nuclear Medicine and Molecular Imaging; Amyloid Imaging Taskforce. Appropriate Use Criteria for Amyloid PET: A Report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *Alzheimers Dement.* 2013 Jan;9(1):e-1-16. doi: 10.1016/j.jalz.2013.01.002.

Johnson JK, Head E, Kim R, Starr A, Cotman CW. Clinical and pathological evidence for a frontal variant of Alzheimer disease. *Arch Neurol* 1999;56:1233–9.

Leuzy A, Savitcheva I, Chiotis K, Lilja J, Andersen P, Bogdanovic N, Jelic V, Nordberg A. Clinical impact of [18F]flutemetamol PET among memory clinic patients with an unclear diagnosis. *Eur J Nucl Med Mol Imaging.* 2019 Jun;46(6):1276-1286. doi: 10.1007/s00259-019-04297-5. Epub 2019 Mar 26. PMID: 30915522; PMCID: PMC6486908.

Lijmer JG, Mariska Leeflang and Patrick M. M. Bossuyt. Proposals for a Phased Evaluation of Medical Tests. *Med Decis Making* (2009a) 29: E13 originally published online 15 July 2009

Lijmer JG, Bossuyt PM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol.* 2009 Apr;62(4):364-73. doi: 10.1016/j.jclinepi.2008.06.017. Epub 2008 Oct 21. PMID: 18945590.

- Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (SteeleRichardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. *Neurology*. 1996;47:1–9
- McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology*. 2005;65:1863–72.
- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH. 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. *Alzheimers Dement*. 2011 May;7(3):263-9. doi: 10.1016/j.jalz.2011.03.005. Epub 2011 Apr 21. PMID: 21514250; PMCID: PMC3312024.
- Ossenkuppele R, Prins ND, Pijnenburg YA, Lemstra AW, van der Flier WM, Adriaanse SF, Windhorst AD, Handels RL, Wolfs CA, Aalten P, Verhey FR, Verbeek MM, van Buchem MA, Hoekstra OS, Lammertsma AA, Scheltens P, van Berckel BN. Impact of molecular imaging on the diagnostic process in a memory clinic. *Alzheimers Dement*. 2013 Jul;9(4):414-21. doi: 10.1016/j.jalz.2012.07.003. Epub 2012 Nov 16. PMID: 23164552.
- Paulsen JS, Nance M, Kim J-I, Carlozzi NE, Panegyres PK, Erwin C, Goh A, McCusker E, Williams JK (2013) A review of quality of life after predictive testing for and earlier identification of neurodegenerative diseases. *Prog Neurobiol* 110, 2-28.
- Rabinovici GD^{1,2}, Gatsonis C^{3,4}, Apgar C⁵, Chaudhary K¹, Gareen I^{3,6}, Hanna L³, Hendrix J⁷, Hillner BE⁸, Olson C⁵, Lesman-Segev OH¹, Romanoff J³, Siegel BA⁹, Whitmer RA^{10,11}, Carrillo MC⁷. Association of Amyloid Positron Emission Tomography With Subsequent Change in Clinical Management Among Medicare Beneficiaries With Mild Cognitive Impairment or Dementia. *JAMA*. 2019 Apr 2;321(13):1286-1294. doi: 10.1001/jama.2019.2000.
- Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*. 2011;134:2456–77.
- Roman GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*. 1993;43:250–60.
- Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess*. 2007 Dec;11(50):iii, ix-51. doi: 10.3310/hta11500. PMID: 18021577.
- Sanchez-Juan et al. Practical utility of amyloid and FDG-PET in an academic dementia center. *Neurology* 82 January 21, 2014
- Surti S, Kuhn A, Werner ME, Perkins AE, Kolthammer J, Karp JS. Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. *J Nucl Med*. 2007;48:471–80.
- Triviño-Ibáñez EM, Sánchez-Vañó R, Sopena-Novales P, Romero-Fábrega JC, Rodríguez-Fernández A, Carnero Pardo C, Martínez Lozano MD, Gómez-Río M. Impact of amyloid-PET in daily clinical management of patients with cognitive impairment fulfilling appropriate

use criteria. *Medicine (Baltimore)*. 2019 Jul;98(29):e16509. doi: 10.1097/MD.00000000000016509. PMID: 31335725; PMCID: PMC6708756.

Vos et al. Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *BRAIN* 2015: 138; 1327–1338

Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004 Feb 3;140(3):189-202. doi: 10.7326/0003-4819-140-3-200402030-00010. PMID: 14757617.

Zwan MD, Bouwman FH, Konijnenberg E, van der Flier WM, Lammertsma AA, Verhey FR, Aalten P, van Berckel BN, Scheltens P. Diagnostic impact of [18F]flutemetamol PET in early-onset dementia. *Alzheimers Res Ther*. 2017 Jan 17;9(1):2. doi: 10.1186/s13195-016-0228-4. PMID: 28093088; PMCID: PMC5240413.

TABLES / GRAPHS

Table 1: Sample characteristics at baseline by post-PET diagnosis (mean, SD or n, %).

	AD (n=106)	FTD (n=18)	other NDD (n=18)	no NDD (n=10)
Age	62 (6.1)	62 (4.7)	64 (5.3)	60 (5.8)
Female gender	51 (48%)	7 (39%)	6 (33%)	3 (30%)
MMSE	23 (3.2)	24 (3.3)	25 (4.3)	25 (4.2)
DAD	85 (14.8)	80 (21.2)	87 (12.3)	88 (14.8)
GDS	3 (2.3)	3 (2.4)	4 (2.9)	3 (2.2)
post-PET diagnosis				
AD	86 (81%)	3 (17%)	2 (11%)	5 (50%)
VAD	1 (1%)		3 (17%)	
FTD	4 (4%)	15 (83%)		
DLB	2 (2%)		6 (33%)	
other NDD	3 (3%)		7 (39%)	
no NDD	10 (9%)			5 (50%)

Abbreviations: AD, Alzheimer's disease; DAD, disability assessment dementia (scale range 0-100, higher is less affected); DLB, dementia with Lewy bodies; FTD, frontotemporal dementia; GDS, geriatric depression scale (scale range 0-15, higher is more affected); MMSE, Mini-Mental State Examination (scale range 0-30, higher is less affected); NDD, neurodegenerative disease; VAD, vascular dementia.

Table 2: pre-PET diagnosis versus post-PET diagnosis, stratified by the reference diagnosis.

Pre-PET	Post-PET	Reference diagnosis	Pre/post-PET change?	If change, in line with reference?	Regardless of change, post-PET in line with reference?	Comment	n ¹	n ²	
<i>CORRECT CHANGE</i>									
noAD	AD	AD	Yes	Yes	Yes		5	7	A
AD	noAD	noAD	Yes	Yes	Yes		3	3	B
AD	noAD	mixed noAD	Yes		Yes	post-PET within mixed	1	1	C
<i>INCORRECT CHANGE</i>									
AD	noAD	AD	Yes	No	No		4	5	D
AD	noAD	mixed AD	Yes	No	No	post-PET outside mixed	1	2	E
AD	noAD	noAD	Yes	No	No	post-PET different noAD than reference noAD	2	4	F
<i>CHANGE WITHIN MIXED</i>									
AD	noAD	mixed AD	Yes		Yes	post-PET within mixed	4	5	H
<i>NO CHANGE (INCORRECT)</i>									
AD	AD	noAD	No	n/a	No		2	3	N
noAD	Same noAD	AD	No	n/a	No		2	3	O
noAD	Same noAD	noAD	No	n/a	No	post-PET different noAD than reference noAD	4	5	P
<i>NO CHANGE (CORRECT)</i>									
AD	AD	AD	No	n/a	Yes		65	70	I
AD	AD	mixed AD	No	n/a	Yes		11	12	J
noAD	same noAD	noAD	No	n/a	Yes		13	18	K
noAD	same noAD	mixed AD	No	n/a	Yes	post-PET within mixed	3	4	L
noAD	same noAD	mixed noAD	No	n/a	Yes	post-PET within mixed	2	2	M
<i>NO CONSENSUS</i>									
AD	AD	no consensus	n/a	n/a	n/a		8	1	Q
AD	noAD	no consensus	n/a	n/a	n/a		5	1	R
noAD	AD	no consensus	n/a	n/a	n/a		5	2	S
noAD	same noAD	no consensus	n/a	n/a	n/a		11	3	T
noAD	other noAD	no consensus	n/a	n/a	n/a		1	1	U
<i>Total</i>							152	152	

¹ Excluding plausible evident cases.

² Including plausible evident cases.

[Alternative table for consideration]

Pre-PET	Post-PET	Change in clinical diagnosis after PET?	correct change according to reference standard	incorrect change according to reference standard	change within mixed diagnosis set in reference standard	no change (diagnosis different from reference diagnosis)	no change (diagnosis same as reference diagnosis)	no consensus
noAD	AD	Yes	5					5
AD	noAD	Yes	4	7	4			5
AD	AD	No				2	76	8
noAD	same noAD	No				6	18	11
noAD	other noAD	Yes						1

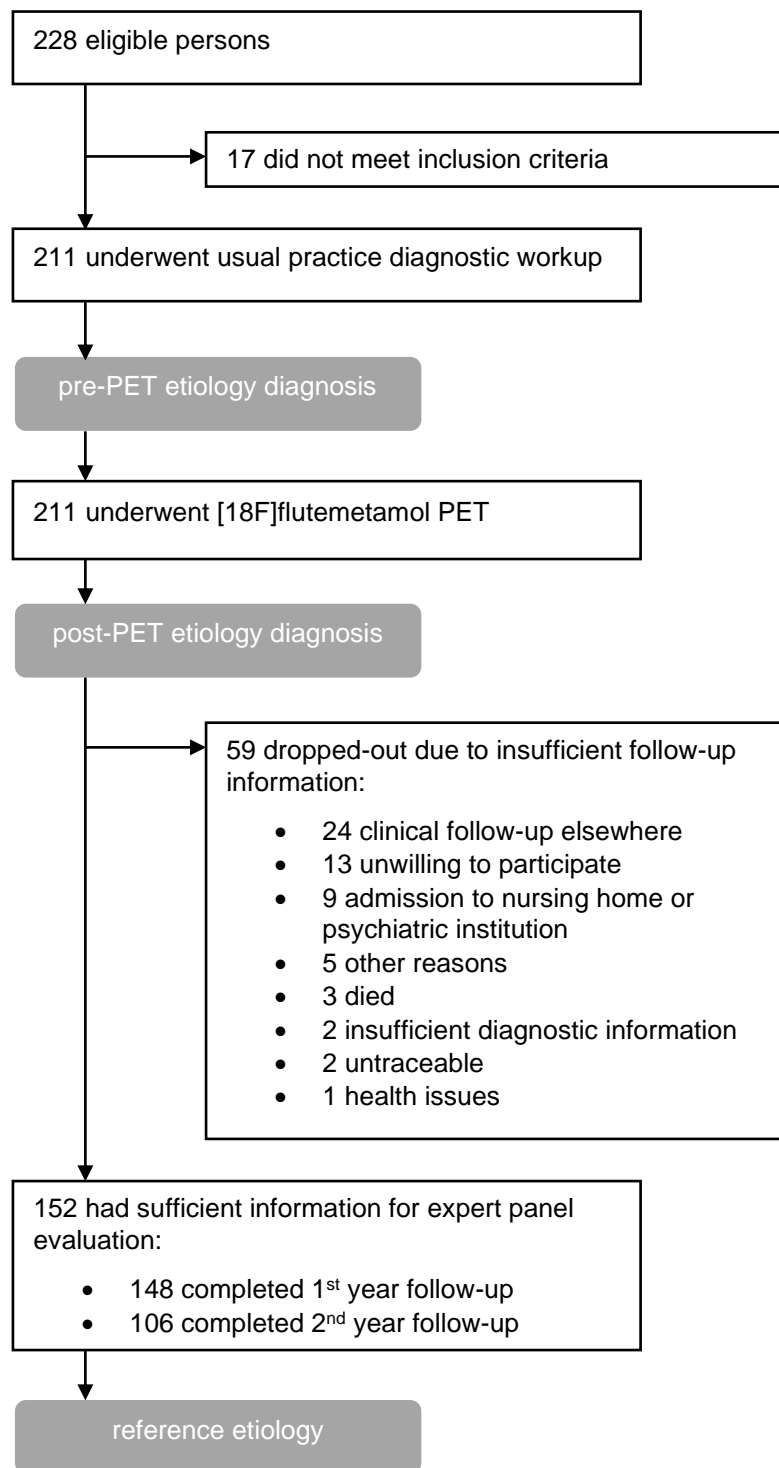


Figure 1: Flowchart of patient inclusion, follow-up and diagnostic accuracy design.

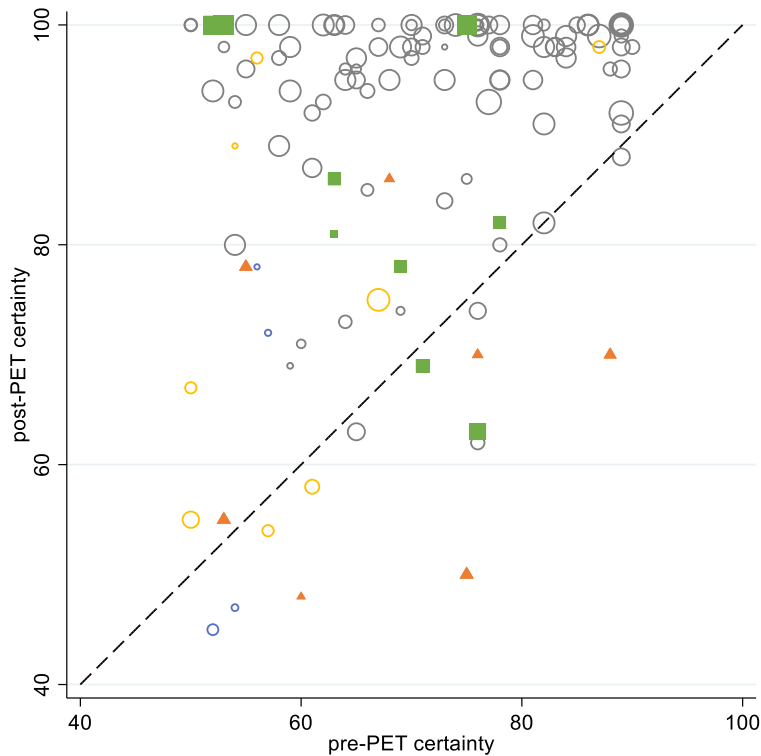


Figure 2: Pre- and post-PET certainty of each participant classified according to table 2: green filled squares: change concordant with reference standard, orange filled triangles: change discordant with reference standard, blue open circles: changed within mixed, yellow open circles: no change and concordant to reference standard, gray open circles: no change and concordant to reference standard; dashed black line is diagonal; circle size is reflection of the relative certainty within each of the 5 classifications.

SUPPLEMENTAL 1: IN- AND EXCLUSION CRITERIA

INCLUSION CRITERIA

Patients:

- Who are aged less than 70 at the moment of the diagnosis;
- Who have dementia or are suspected of dementia;
- Whose etiological diagnosis was set with a certainty between 50-90%;
- Who underwent a brain MRI during a standardized screening;
- Who signed an informed consent prior to the study performances;
- Who are mentally competent (this was operationalized by a Mini Mental State Examination score ≥ 18);
- Whose body weight is more or equal to 50 kilos (this was done due to PET scanning properties).

EXCLUSION CRITERIA

Patients:

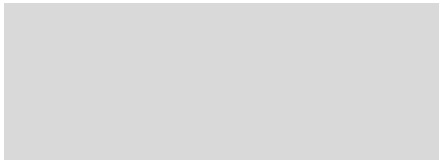
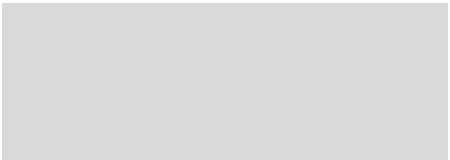
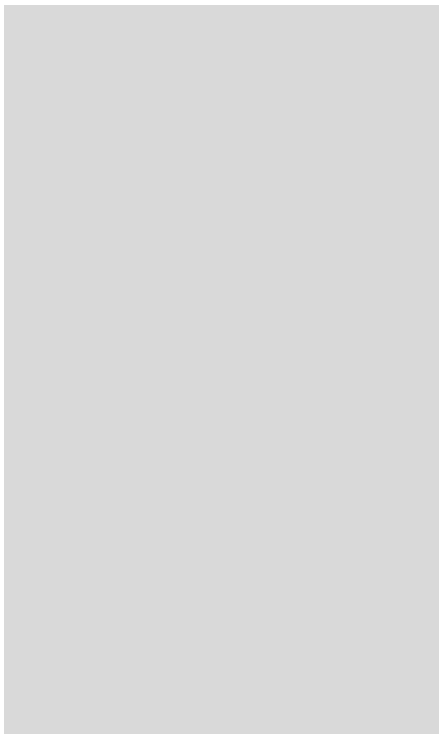
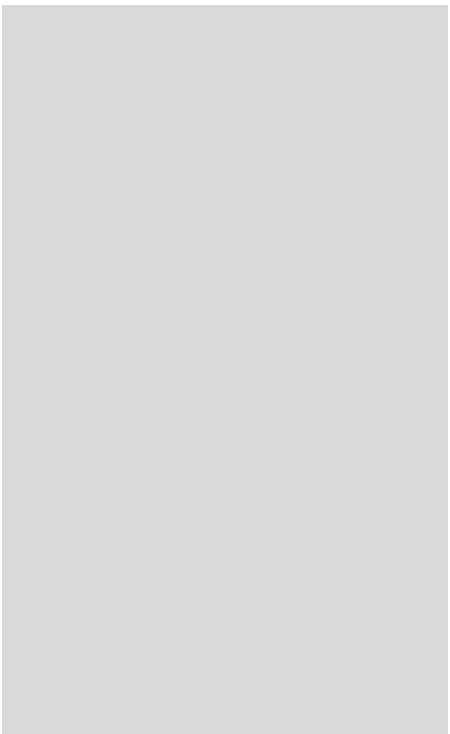
- Who are considered medically unstable;
- Who require additional laboratory tests or workup between enrolment and completion of the [18F]flutemetamol PET scan;
- Who have a clinically significant infectious disease, including Acquired Immunodeficiency Syndrome (AIDS) or Human Immunodeficiency Virus (HIV) infection;
- Who are receiving any investigational medications, or have participated in a trial with investigational medications within the last 30 days;
- Who have ever participated in an experimental study with an amyloid targeting agent (e.g. anti-amyloid immunotherapy, γ -secretase or γ -secretase inhibitor) unless it can be documented that the subject received only placebo during the course of the trial;
- Who have had a radiopharmaceutical imaging or treatment procedure within 7 days prior to the study imaging session;
- Who are females of childbearing potential who are not surgically sterile, not refraining from sexual activity or not using reliable methods of contraception. Females of childbearing potential must not be pregnant (negative serum β -hCG at the time of screening and negative urine β -hCG on the day of imaging) or breast feeding at screening. Females must avoid becoming pregnant, and must agree to refrain from sexual activity or to use reliable contraceptive methods such as prescribed birth control or IUD for 24 hours following administration of [18F]flutemetamol;
- Who are claustrophobic;
- Who have abnormalities on MRI other than white matter changes or an incidental small lacunar lesion;
- Who have metal objects in or around the body (braces, pacemaker, metal fragments).

SUPPLEMENTAL 2: EXAMPLE VIGNET WITH HYPOTHETICAL DIAGNOSTIC INFORMATION

	Baseline	1-year follow-up	2-year follow-up
Age	77		
Gender	female		
Education	Higher Professional Education		
Clinical history	Patient indicates that everything requires a bit more time before things get through to her. Exercise is difficult due to her stiff muscles. She indicates her memory is good and concentration during reading is good if there is sufficient interest in the topic. She noticed to lose stuff more often.	Patient indicates to have trouble hearing. She does not indicate a memory decline.	Patient indicates that memory has declined. Sleeping is OK but she lost appetite. She can't name her daily activities.
Informant-based history	According to her husband memory and orientation in time has declined. When providing clues she remembers things.	Husband reports further decline of memory, especially short-term memory. Furthermore, he things she is disoriented and she does not ask for help by asking things again if she did not hear it well enough.	According to husband and son physical functioning has declined. Orientation is difficult. Husband has to help the patient with many things.
One or more 1st degree relatives with dementia	No		
Mini-mental State Examination (MMSE)	30	26	22
Clinical Dementia Rating scale (CDR)	0.5	0.5	1
Neuropsychiatric inventory (NPI)	0/144	0/144	0/144
Geriatric depression scale (GDS-15)	1	0	1
Disability assessment for Dementia (DAD)	100% initiative 100% planning & organisation 94% effective implementation	98% initiative 100% planning & organisation 91% effective implementation	84% initiative 92% planning & organisation 76% effective implementation
DAD (total score)	98%	96%	84%
BMI	22.4	22.2	22.6
Systolic / diastolic blood pressure	92 / 128	96 / 173	91 / 159
General physical examination	Normal (no particularities)		
General neurological examination	Some rigidity in both arms, loss of amplitude at finger tapping. Gesture and action tremor.		
Balance disorder	No	No	No
Pyramidal Symptoms	No	No	No
Lateralisation	No	No	No
Walking difficulties	No	No	No
canine paresis	No	No	No
Incontinence	No	No	No
Rigidity	Yes	Yes	Yes
tremor	Yes	Yes	Yes
bradykinesia	No	No	No
Somatic conditions	Coronary bypass (2009), diabetes	Coronary bypass, diabetes, hearing problems	Coronary bypass, diabetes
Smoking intake	Quite smoking		
Alcohol consumption per week	1	[unknown]	[unknown]
Medication use	Acetylcardio, ascal, metformine	Acetylcardio, ascal, metformine	Acetylcardio, ascal, metformine

	RAW SCORE	Z-SCORE	INTERPRET		RAW SCORE	Z-SCORE	INTERPRET		RAW SCORE	Z-SCORE	INTERPRET
15-Word Learning Task (WLT) total	1+3+4+ 5+6=25	-2.5	Very low		3+4+6+ 6+7=26	-1.6	Low		3+4+4+ 5+3=19	-2.4	Very low
15-WLT delayed recall	4	-1.6	Low		6	-0.60	Below average		3	-1.9	Low
15-WLT retention (max. score)	4 / 6	-0.17	Average		6 / 7	0.88	Above average		0 / 6	-3.7	Very low
Visual Association Test trial 1	4				1				[unknown]		
Visual Association Test trial 2	6				4				[unknown]		
Visual Association Test Fluency	10	-1.35	Low		5	-2.35	Very low		[unknown]	[unknown]	[unknown]
Digit-span	25	0.05	Average		11	-2	Very low		12	-1.7	Low
Visual Object and Space Perception Battery	8 / 4		Average		8 / 3		Average		7 / 4		Below average
Letter Digit Substitution Test (60 seconds)	18 / 9				19 / 10		Below average		20 / 10		
Stroop Color-Word Test 1	18	-1.5	Low		21	-1.4	average		8	-3.2	Very low
Stroop Color-Word Test 2	58	-1.6	Low		56	-1.6	Low		67	-3.2	Very low
Stroop Color-Word Test 3	88	-2.5	Very low		87	-2.4	Very low		105	-3.7	Very low
Stroop interference (3-((2+1)/2))	181	-2.4	Very low		242	-4.9	Very low		241	-4.9	Very low
Trail Making Test A	112	-2.1	Very low		167	-4.6	Very low		192	-4.7	Very low
Trail Making Test B	91	-2.2	Very low		106	-2.5	Very low		109	-2.5	Very low
Trail Making Test B/A index	141	-0.64	Below average		212	-1.7	Low		[unknown]	[unknown]	[unknown]
Conclusion neuropsychological examination	1.52				2.01				[unknown]		
	The neuropsychological examination indicates a low mental speed. Memory performance is average except active information uptake.				The patient scores mainly low on tests of mental speed, attention/concentration and executive functioning. Compared to last year several domains declined.				Scores on mental speed are very low though stable compared to last year. Memory declined except recognition.		

Medial temporal lobe atrophy score left (0-4)	1
Medial temporal lobe atrophy score right (0-4)	2
Global cortical atrophy (0-3)	1
Parietal atrophy left (0-3) Koedam score	2
Parietal atrophy right (0-3) Koedam score	1
White matter lesions (Fazekas) (0-3)	1
Infarcts	0
Lacunae	0
Microbleeds	0
MRI summary	Normal and symmetrical aspect of ventricular system and peripheral CSFs with good gray/white matter separation. Minor global cortical atrophy (grade I). There is hippocampal atrophy grade II right, grade I left. No watershed other than thalamic infarcts. No space-consuming process. No microbleeds detectable. Conclusion: Neurodegenerative features with pathological hippocampal atrophy on the right as described above, probably in the context of Alzheimer's pathology.
Amyloid beta 42 (pg/ml) (<640 is abnormal)	1193
Total tau (pg/ml) (> 375 is abnormal)	283
Phosphorylated-tau (pg/ml) (> 52 is abnormal)	37
APOE	E3E3
Additional investigations	FDG-PET scan shows reduced mesial frontal uptake extending to lateral frontal both sides. Elsewhere normal uptake in temporal, parietal, occipital, basal nuclei, thalamus, mesencephalon and cerebellum. The Alzheimer's discrimination tool is negative. Conclusion: Subtly deviating FDG-PET scan, possible early frontotemporal dementia.



SUPPLEMENTAL 3: DIAGNOSTIC ASSESSMENTS

PRE-PET DIAGNOSIS

Item 1: Most likely etiology (choose 1 option):

- Alzheimer
- Vascular
- Frontotemporal
- Lewy Bodies
- Other neurodegenerative condition, namely _____
- No neurodegenerative condition, namely _____
- Unclear / postponed

Item 2: Degree of certainty on most likely etiology (mark on the line):

Completely uncertain

Completely certain

|-----|

Item 3: Etiological differential diagnosis (choose 1 option):

- Alzheimer
- Vascular
- Frontotemporal
- Lewy Bodies
- Other neurodegenerative condition, namely _____
- No neurodegenerative condition, namely _____

POST-PET DIAGNOSIS

Item 1: Most likely etiology (choose 1 option):

- Alzheimer
- Vascular
- Frontotemporal
- Lewy Bodies
- Other neurodegenerative condition, namely _____
- No neurodegenerative condition, namely _____
- Unclear / postponed

Item 2: Degree of certainty on most likely etiology (mark on the line):

Completely uncertain

Completely certain

|-----|

SUPPLEMENTAL 4: EXAMPLES OF MAJORITY

EXAMPLE 1

Individual expert's scores (on scale of 1 to 15):

Expert 1: 7 for AD and 3 for VAD.

Expert 2: 11 for AD and 7 for VAD.

Expert 3: 9 for AD, 4 for VAD and 2 for other NDD.

The weighted scores are:

Expert 1: AD: $7/(7+3) = 0.70$; VAD: $3/(7+3) = 0.30$

Expert 2: AD: $11/(11+7) = 0.61$; VAD: $7/(11+7) = 0.39$

Expert 3: AD = $10/(10+3+2) = 0.67$; VAD: $4/(10+3+2) = 0.20$; other NDD $2/(10+3+2) = 0.13$

Diagnoses with weight higher than or equal to 0.33:

Expert 1: AD

Expert 2: AD, VAD

Expert 3: AD

Majority diagnosis:

AD

EXAMPLE 2

Individual expert's scores:

Expert 1: 7 for AD and 3 for VAD (on scale of 1 to 15).

Expert 2: 11 for AD and 7 for VAD.

Expert 3: 9 for AD and 4 for VAD.

The weighted scores are:

Expert 1: AD: $7/(7+3) = 0.70$, VAD: $3/(7+3) = 0.30$

Expert 2: AD: $11/(11+7) = 0.61$, VAD: $7/(11+7) = 0.39$

Expert 3: AD = $9/(9+4) = 0.69$, VAD: $4/(9+4) = 0.31$

Diagnoses with weight higher than or equal to 0.33:

Expert 1: AD

Expert 2: AD, VAD

Expert 3: AD

Majority diagnosis:

AD

SUPPLEMENTAL 5: NO (PLAUSIBLE) MAJORITY

Ratings from 3 experts on etiology (“Which etiology determines the clinical picture”; scale from 1 to 15) for those 8 participants in which the no majority diagnosis was reached on the etiology and no plausible evident etiology could be set.

participant	expert	AD	VAD	FTD	DLB	Other	certainty	pre-PET	post-PET
A	1	11	1	1	1	1	7	FTD	AD
	2	12	1	12	1	1	10		
	3	2	1	14	1	1	14		
B	1	7	1	1	1	6	4	no NDD	no NDD
	2	9	1	6	4	14	9		
	3	5	1	1	1	8	4		
C	1	8	1	1	1	1	8	AD	no NDD
	2	10	1	2	2	12	5		
	3	3	1	1	1	9	5		
D	1	12	1	11	7	7	9	FTD	FTD
	2	2	1	10	2	1	11		
	3	4	1	9	12	12	12		
E	1	13	1	1	1	9	12	AD	AD
	2	11	1	2	4	3	6		
	3	8	1	13	1	1	13		
F	1	13	1	1	1	3	11	no NDD	AD
	2	7	1	1	7	1	4		
	3	1	1	1	1	15	11		
G	1	1	5	1	1	15	11	other NDD	other NDD
	2	2	3	2	10	3	5		
	3	6	10	3	9	10	8		
H	1	13	1	1	1	5	11	other NDD	DLB
	2	2	1	1	13	1	12		
	3	9	6	3	7	12	9		