

^{18}F -FDG PET for Prediction of Conversion to Alzheimer's Disease Dementia in People with Mild Cognitive Impairment: An Updated Systematic Review of Test Accuracy

Nadja Smailagic^{a,1,*}, Louise Lafortune^{a,1}, Sarah Kelly^a, Chris Hyde^b and Carol Brayne^a

^a*Cambridge Institute of Public Health, Forvie Site, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, UK*

^b*Exeter Test Group and South West CLAHRC, University of Exeter Medical School, St Luke's Campus, Exeter, UK*

Accepted 30 May 2018

Abstract.

Background: A previous Cochrane systematic review concluded there is insufficient evidence to support the routine use of ^{18}F -FDG PET in clinical practice in people with mild cognitive impairment (MCI).

Objectives: To update the evidence and reassess the accuracy of ^{18}F -FDG-PET for detecting people with MCI at baseline who would clinically convert to Alzheimer's disease (AD) dementia at follow-up.

Methods: A systematic review including comprehensive search of electronic databases from January 2013 to July 2017, to update original searches (1999 to 2013). All key review steps, including quality assessment using QUADAS 2, were performed independently and blindly by two review authors. Meta-analysis could not be conducted due to heterogeneity across studies.

Results: When all included studies were examined across all semi-quantitative and quantitative metrics, exploratory analysis for conversion of MCI to AD dementia ($n = 24$) showed highly variable accuracy; half the studies failed to meet four or more of the seven sets of QUADAS 2 criteria. Variable accuracy for all metrics was also found across eleven newly included studies published in the last 5 years (range: sensitivity 56–100%, specificity 24–100%). The most consistently high sensitivity and specificity values (approximately $\geq 80\%$) were reported for the sc-SPM (single case statistical parametric mapping) metric in 6 out of 8 studies.

Conclusion: Systematic and comprehensive assessment of studies of ^{18}F -FDG-PET for prediction of conversion from MCI to AD dementia reveals many studies have methodological limitations according to Cochrane diagnostic test accuracy gold standards, and shows accuracy remains highly variable, including in the most recent studies. There is some evidence, however, of higher and more consistent accuracy in studies using computer aided metrics, such as sc-SPM, in specialized clinical settings. Robust, methodologically sound prospective longitudinal cohort studies with long (≥ 5 years) follow-up, larger consecutive samples, and defined baseline threshold(s) are needed to test these promising results. Further evidence of the clinical validity and utility of ^{18}F -FDG PET in people with MCI is needed.

Keywords: Accuracy, Alzheimer's disease dementia, conversion, ^{18}F -FDG PET, mild cognitive impairment, test predictive value

¹These authors contributed equally to this work.

*Correspondence to: Nadja Smailagic, Cambridge Institute of Public Health, Forvie Site, University of Cambridge School of

Clinical Medicine, Box 113, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK. Tel.: +44 1223 767149; E-mail: ns530@medschl.cam.ac.uk.

INTRODUCTION

The diagnosis of probable Alzheimer's disease (AD) has classically been based on clinical criteria and postmortem confirmation of AD [1]. The reconceptualization of AD as a disease continuum [2]—with a long asymptomatic phase followed by a symptomatic phase of progressive cognitive decline before the onset of functional impairment and overt dementia—has led to a shift towards identifying AD at an earlier stage, before patients have crossed the threshold into dementia [3].

Clinical subtypes of mild cognitive impairment (MCI) have been regarded as prodromal forms of a variety of dementias [4]. There are four outcomes for patients with MCI: progression to AD dementia, progression to another dementia type, maintaining stable MCI, or recovery. Studies [5–8] indicate that an annual average of 10% to 15% of people with MCI progress to AD dementia. Discriminating between patients who will and will not progress to dementia due to AD is critical in the context of care and future therapies [9].

¹⁸F-FDG PET provides information about neuronal activity at tissue level by measuring regional cerebral glucose metabolic rate (CGMr). Glucose hypometabolism in the temporo-parietal lobe and posterior cingulate cortex, as assessed by ¹⁸F-FDG PET, signals a pattern of neuronal loss and synaptic dysfunction typically found in AD dementia [10, 11]. ¹⁸F-FDG PET is a valued biomarker test in the early, confirmatory diagnostic arsenal of AD dementia [12, 13].

For MCI, the ¹⁸F-FDG PET pattern is not consistent and usually presents as mild global and regional hypometabolism in patients diagnosed with MCI in research settings [14]. Nevertheless, several ¹⁸F-FDG PET studies in MCI patients have found characteristic and progressive CGMr reductions in particular AD-vulnerable regions, suggesting that certain findings on brain PET scans can potentially predict progression of MCI to AD dementia [15–18].

A previous Cochrane review (published in 2014 based on searches to 2013) [19] concluded that the evidence from studies ($n = 16$) published over a 14-year period shows variability in prognostic performance of ¹⁸F-FDG PET, and was insufficient to support the routine use of ¹⁸F-FDG PET in clinical practice in people with MCI. This conclusion was challenged [20] in light of further published studies and some concerns were expressed regarding study design, data analysis procedures, and lack of adequate

and validated image analysis approaches in included studies.

This paper fully updates the original review, including studies published over the last five years, and maintains the same robust methodological approach. The main objectives are to: 1) reassess the accuracy and reliability of ¹⁸F-FDG-PET for conversion from MCI to AD dementia in light of the quality and quantity of the most recent evidence; 2) include an assessment of prognostic value among different metrics; 3) provide research recommendations to further strengthen the evidence base and move towards routine clinical utility.

MATERIALS AND METHODS

Search strategy and selection criteria

This is a systematic review of diagnostic test accuracy (DTA) studies that applied the same search strategy and methodology as the original Cochrane ¹⁸F-FDG-PET review [19], unless otherwise specified. Methods followed those of the Cochrane Handbook for DTA studies (<http://methods.cochrane.org/sdt/handbook-dta-reviews>).

Searches were conducted of electronic databases from January 2013 to July 2017 to update the original Cochrane review [19] (Supplementary Table 1). No language restrictions or search filters were applied [21]. Two review authors (NS, LL) independently screened references and examined reference lists of any relevant studies and systematic reviews to identify additional studies, and independently extracted data. Differences were resolved by discussion.

Study design

Prospective longitudinal, nested case-control cohort studies and cohorts that analysed data retrospectively were eligible if they contained sufficient data to construct two-by-two tables expressing ¹⁸F-FDG-PET results by disease status.

Both short and long periods of observation are needed to understand the role of biomarkers in predicting cognitive decline in people with MCI [22] and assess their accuracy. Because a 'short' follow-up interval has been defined as at least two years by some [23–25], and as one–three years by others [26], this review included studies with one year as the minimum period of follow-up for diagnostic verification. Cross-study variability in test accuracy according to

follow-up intervals could not be examined due to insufficient number of studies.

Participants

Participants with MCI recruited from any setting were eligible if studies used the Petersen criteria [4, 27], or any of the classifications to describe MCI syndrome in Matthews et al. [28]. These broad definitions of MCI were applied in order not to exclude any study based only on criteria for MCI, and to assess the accuracy of ¹⁸F-FDG scans in identifying a feature of AD pathology in people with cognitive impairment [29].

Analytical method/metric for ¹⁸F-FDG PET

Which standardization approaches (i.e., semi-quantification methods, observer-independent analyses, identification of cut-off values) are appropriate for best predictive value of ¹⁸F-FDG PET at MCI stage are still a matter of debate [20]. Recent literature [13–30] suggests that a quantitative approach is preferable and describes a number of tools that are widely used for automated FDG PET analysis, for instance: 1) Statistical Parametric Mapping (SPM); 2) three-dimensional Stereotactic Surface Projection (3-D-SSP) statistics (Neurostat); 3) AD t-sum implemented in PMOD, etc. However, there are currently neither generally accepted standards to define a ¹⁸F-FDG abnormality threshold nor a preferred and widely accepted metric; therefore, studies are included irrespective of analytical method.

Target condition/outcome

Rates of conversion were based on clinical assessment. The target condition was conversion to clinically diagnosed AD dementia. Studies that applied the probable or possible National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria [1] or other widely used clinical criteria for AD dementia [31, 32] are included.

Exclusion criteria

Studies were excluded if they focused on people with a secondary cause for cognitive impairment, namely: 1) current use or history of alcohol/drug abuse; 2) central nervous system trauma, tumor,

or infection; 3) other neurological conditions, e.g., Parkinson's or Huntington's diseases.

Extraction of data from individual studies

Data was extracted on study characteristics, PET acquisition, analytical approach, data for two-by-two tables, and methodological quality. Some study authors were contacted to obtain data for creating two-by-two tables and/or missing data and/or data for quality assessment. RevMan 5.3 software (Cochrane Collaboration) [33] was used for data collection and management.

Data analysis

The accuracy of ¹⁸F-FDG was evaluated according to the target condition. Data in the two-by-two tables (binary test results cross-classified with the binary reference standard) (Cochrane Handbook, <http://met.hods.cochrane.org/sdt/handbook-dta-reviews>) were used to calculate sensitivities and specificities, with 95% confidence intervals. Exploratory analyses were conducted by plotting estimates of sensitivity and specificity from each study on forest plots. As included studies used a wide range of thresholds, it was inappropriate to meta-analyse pairs of sensitivity and specificity using a bivariate random-effect approach [34], as first intended. Instead, a summary of exploratory analyses is provided.

The intention was to investigate, using subgroup and sensitivity analyses, the effect of 1) analytical approach (metric), 2) pre-specification of threshold, and 3) study design on the diagnostic accuracy of the ¹⁸F-FDG PET index test. The plan was to also explore the impact of 1) duration of follow-up, 2) MCI diagnostic criteria, 3) clinical setting, and 4) type of reference standard on the summary estimates. However, for the reasons described above, a meta-analysis was not feasible. Had a meta-analysis been possible, our plan had been to perform meta-regression by including each potential source of heterogeneity as a covariate in order to formally assess their effects on any summary accuracy estimates.

Quality assessment

Methodological quality was assessed using Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) [35] independently by two review authors (NS, SK) resolving disagreement by further

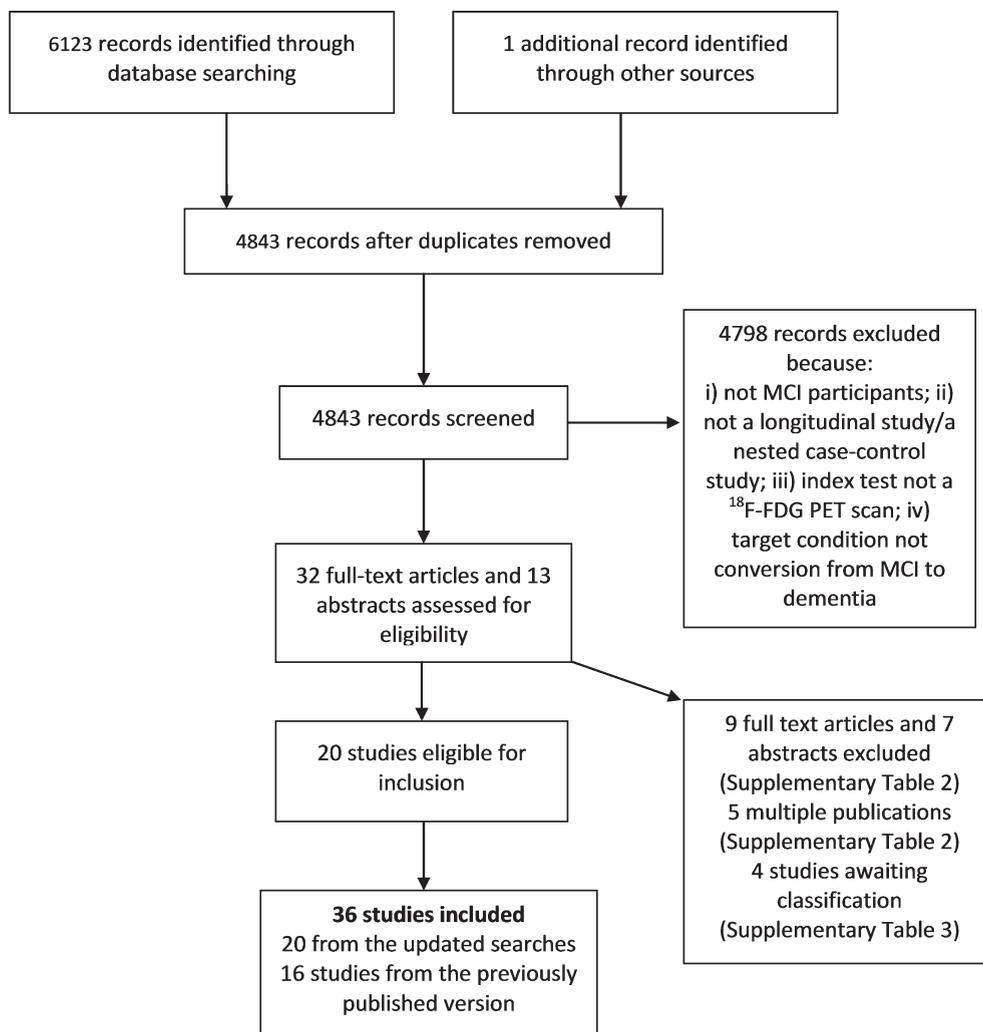


Fig. 1. Study selection from the updated search.

review and discussion, alongside an arbitrating third reviewer, when necessary. The QUADAS 2 tool is made up of seven criteria organized around four domains: 1) Participant selection; 2) Index test; 3) Reference standard; 4) Participant flow. Full details of the questions to assess these domains for risk of bias are presented in Supplementary Table 4.

RESULTS

Search results

The flow chart in Fig. 1 shows the updated searches. We initially identified 45 records for inclusion. Upon further assessment and correspondence with most study authors, 20 studies [24, 36–54] met

the inclusion criteria. Reasons for exclusion are supplied in Fig. 1 and Supplementary Tables 2 and 3. No extra studies were identified through reference checking.

Assessment of the accuracy of ¹⁸F-FDG-PET for conversion from MCI to AD dementia

We included 36 studies overall: 16 studies [16, 17, 55–68] from the original review [19] and 20 from the updated search, of which 12 (3 ‘old’ and 9 ‘new’) [36, 40, 41, 45, 49, 51, 54, 61, 62, 68] analysed data from participants from the same Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort (Supplementary Tables 4 and 5). To avoid participant double counting only one (new) ADNI study [45] is included in the

Table 1
Accuracy figures of ¹⁸F-FDG PET for conversion from MCI to AD dementia at study level

Metrics (analytic approach/image analysis)	Study ID	Participants (No. of ¹⁸ F-FDG PET positive*)	Age (y) All MCI participants or converters (non-converters)	Accuracy at study level			No. with MCI stable	Duration of follow-up Mean, Median, Range, Maximum (months)	Sources of recruitment (setting)
				Sensitivity (%)	Specificity (%)	No. with MCI converted (%)			
<i>Computer aided visual read</i>									
Neurostat/3D-SSP	Berent 1999	20 (10)	70 ± 5.5	70	70	10 (50)	10	Maximum 36	Cognitive disorders
	Drzeuga 2005	30 (13)	75 ± 4.7 (68 ± 2.0)	92	89	12 (40)	18	Mean 16.0 ± 2.0	University clinic
	Fellgiebel 2007	16 (7)	69.5 ± 7.9 (68.8 ± 10.0)	100	75	4 (25)	12	Mean 19.6 ± 9.0	Research unit
	Pardo 2010	18 (6)	Mean 80 Range 54–83	25	60	8 (44)	10	Maximum 36	University memory clinic
sc-SPM	Grimmer 2016	28 (10)	62 ± 7.3 Range 50–78	56	74	9 (32)	19	Mean 31.2 ± 7.8	Geriatric, Research, Education and Medical Center
	Ito 2015	88 (43)	71 ± 6.6	63	64	41 (46)	47	36	Cognitive disorders
	Anchisi 2005	48 (19)	71 ± 3.9 (65 ± 9.0)	93	82	14 (29)	34	Median 12 Range 12–27	University center
	Chetelat 2003	17 (7)	73 ± 5.1 (67.8 ± 7.0)	100	100	7 (41)	10	Maximum 18	Memory clinics
	Mosconi 2004	37 (4)	71.5 ± 4.0 (66.0 ± 8.0)	38	97	8 (22)	29	Maximum 18 Mean 12.1 ± 0.6	4 outpatient
	Clerici 2009	26 (23)	74 ± 6.9	92	15	13 (50)	13	Maximum 18 for aMCI 37 for snamCI	University clinics
	Iaccarino 2017	30 (11)	63.6 ± 7.8	77	100	14 (47)	16	Median 26.5	Not reported
	Choo 2013	77 (30)	63 ± 8.2 (59 ± 7.5)	77	80	26 (34)	51	Inter-quartile range 30 Mean 44.0 ± 35.4 Range 1.6–162	Not reported
	Perani 2014**	28 (11)	71 ± 5.7	100	74	5 (18)	23	Mean 27.6 ± 4.1	University Centre for Research and Treatment
	Perani 2016**	28 (9)	69 ± 5.5 (68 ± 7.6)	88	90	8 (28)	20	Mean 27.5 ± 10.4	Memory clinic
SUVr images	Hatashita 2013	68 (57)	Range 50–89	93	24	30 (44)	38	Mean 19.2 ± 7.1	Geriatric Medicine Department Neurology Cognitive Disorder Centre

(Continued)

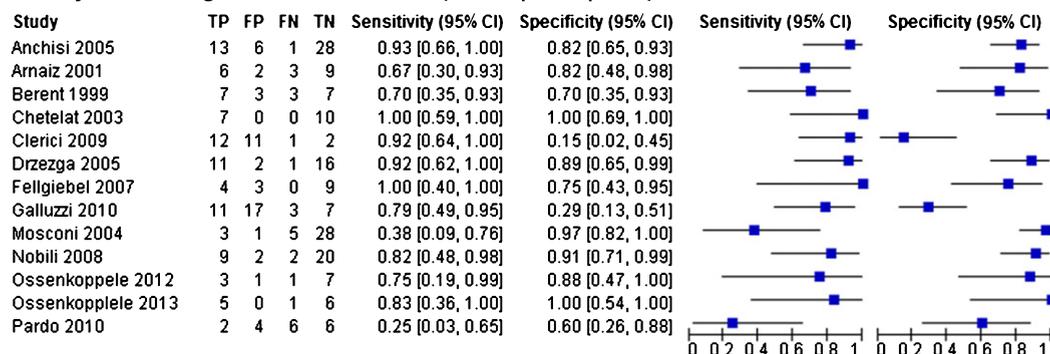
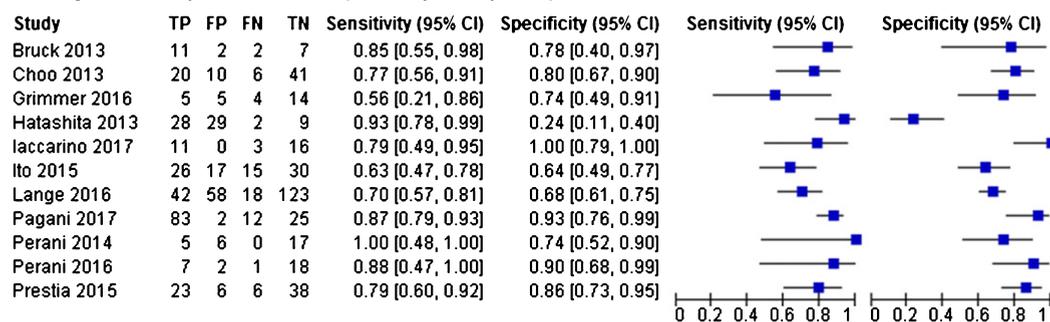
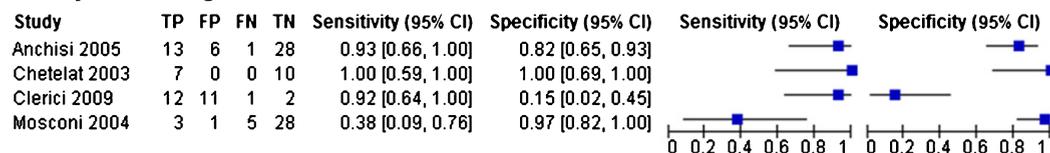
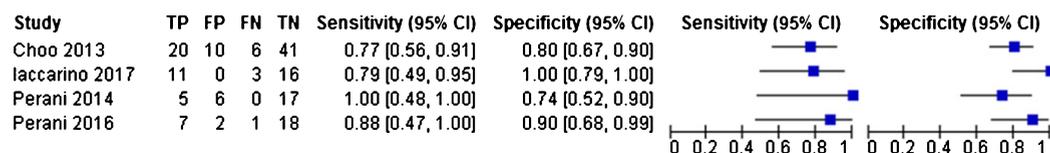
Table 1
(Continued)

Metrics (analytic approach/image analysis)	Study ID	Participants (No. of ¹⁸ F-FDG PET positive*)	Age (y) All MCI participants or converters (non-converters)	Accuracy at study level			Duration of follow-up Mean, Median, Range, Maximum (months)	Sources of recruitment (setting)	
				Sensitivity (%)	Specificity (%)	No. with MCI converted (%)			No. with MCI stable
ROI	Ossenkoppele 2012	12 (4)	67 ± 7.0	75	88	4 (33)	8	Mean 30 Range 24–48 Not reported	Not reported
	Arnaiz 2001	20 (8)	65 ± 8.3 (60 ± 8.4)	67	82	9 (45)	11	Average 36.5 Range 10–75 Maximum 24	Geriatric University clinic
Fully automated read	Bruck 2013	22 (13)	72 ± 7.2 (71 ± 4.9)	85	78	13 (59)	9	Maximum 24	University clinic
	Galluzzi 2010	38 (28)	72.0 ± 7.1	79	29	14 (37)	24	Mean 21.5 ± 10.2	Cognitive disorders University clinic
t-sum/HCI	Grimmer 2016	28 (19)	62 ± 7.3 Range 50–78	78	37	9 (32)	19	Maximum 24	Cognitive disorders University clinic
Principal component analysis / Meta-VOIs	Ito 2015	88 (29)	71 ± 6.6	61	91	41 (46)	47	Maximum 36	University center
	Lange 2016***	241 (100)	74 ± 6.5 (70 ± 7.2)	70	68	60 (25)	181	Maximum 36	Memory clinics ADNI cohort (multicenter) Multicenter
VROI	Prestia 2015	73 (29)	68 ± 8.9 (65 ± 9.4) Range 51–84	79	86	29 (40)	44	Mean 28.0 ± 17.0	Memory clinics ADNI cohort (multicenter) Multicenter
	Nobilit 2008	33 (11)	77 ± 4.8 (75 ± 5.4)	82	91	11 (33)	22	Mean 21.1 ± 10.9	Outpatients Memory clinic
Combined metrics	Pagani 2017	122 (85)		87	93	95 (78)	27	Maximum 60	Memory clinic
	Ossenkoppele 2013	12 (5)	64.0 ± 9.0	83	100	6 (50)	6	Maximum 24	Outpatient University memory clinic

Summary findings of sensitivity and specificity values at study level

Metric	Studies	Converted/Participants	Range of Sensitivity/Specificity	Number of studies with both Sensitivity and Specificity approximately 80% or higher
All metrics	24	450/1132	25%–100% / 15%–100%	14
Computer aided visual read				
Neurostat/3D-SSP	6	84/200	25%–100% / 40%–89%	2
SPM	8	190/413	38%–100% / 15%–100%	6
SUVr/ROI	4	56/122	67%–93% / 24%–88%	2
Fully automated read				
t-sum/HCI	5	153/468	61%–79% / 29%–91%	1
Principal component analysis / Meta-VOIs				
VROI	2	106/155	82%–87% / 91–93	2
Combined metrics				
Combined SUVr images and 't-sum'	1	6/12	83% / 100%	1

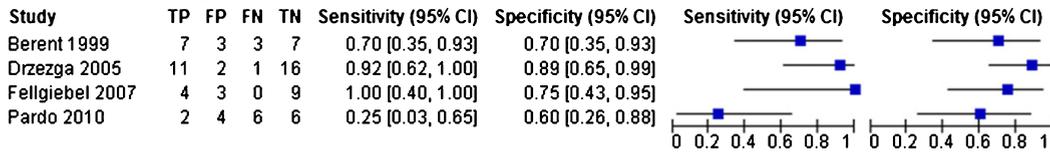
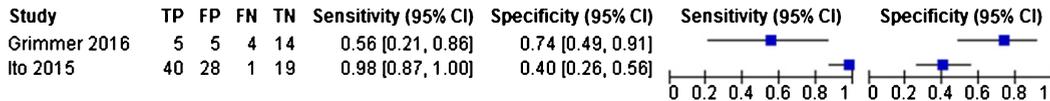
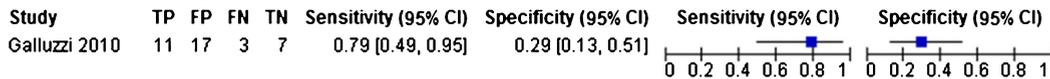
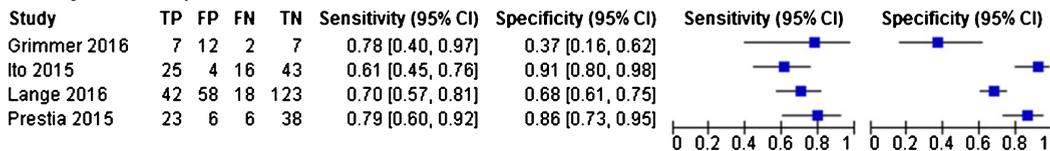
¹⁸F-FDG PET, fluorine-18-2-fluoro-2-deoxy-D-glucose positron emission tomography; MCI, mild cognitive impairment; aMCI, amnesic MCI; snaMCI, single non-amnesic MCI; AD, Alzheimer's disease; 3D-SSP, three-dimensional stereotactic surface projection; sc-SPM, single-case statistical parametric map; SUVr, standardized uptake value ratio; ROI, region of interest; VROI, volumetric region of interest; HCI, hypometabolic convergence index; TOMC, Transitional Outpatient Memory Clinic; SVM, support vector machine. *¹⁸F-FDG PET positive relates to hypometabolism in brain region exceeding a certain threshold. **No overlap between participants in those two studies (Email from Dr. Perani). *** ADNI study, Alzheimer's Disease Neuroimaging Initiative cohort. Notes: All 24 studies used quantitative/semi-quantitative methods. Two studies (Grimmer 2016; Ito 2015) applied two different metrics.

All metrics (N=1132 participants)*Studies from the original Cochrane review (n=327 participants)**Studies from the updated search (n=805 participants)***SPM-single subject metric***Studies from the original Cochrane review**Studies from the updated search*Fig. 2. Forest plots of ¹⁸F-FDG PET for conversion from MCI to Alzheimer's disease dementia.

exploratory analysis for conversion; this study had the largest sample size. One study [38] assessed conversion to AD dementia combined with frontotemporal dementia and dementia with Lewy bodies (LBD). In total, 24 studies were included in the exploratory analysis for conversion to AD dementia (Table 1). Demographic and patient characteristics of MCI participants and information on index test are shown in Supplementary Tables 6 and 7. The majority of studies (21/24) used the Petersen criteria for MCI.

Responses from contacted authors ($n=26$) are presented in Supplementary Table 4.

Individual study estimates of sensitivity and specificity are shown in Figs. 2 and 3 and Table 1. Overall, when studies were examined across all metrics ($n=24$), sensitivity values ranged from 25–100% while specificity values ranged from 15–100%. In the eleven newly identified studies, sensitivity values ranged from 56–100% and specificity from 24–100%. The best values are reported for sc-SPM

Neurostat/3D-SSP metric*Studies from the original Cochrane review**Studies from the updated search***t-sum/HCI metric (fully automated)***Studies from the original Cochrane review**Studies from the updated search*Fig. 3. Forest plots of ¹⁸F-FDG PET for conversion from MCI to Alzheimer's disease dementia.

metric in four new studies; sensitivity values ranged from 77–100% and specificity from 74–100%. In addition to the accuracy of the test for conversion to AD dementia, seven of 24 studies also assessed the accuracy of the test for conversion to AD dementia plus other types of dementia, mainly frontotemporal dementia and LBD (Supplementary Table 8).

Metrics

All 24 studies included in the exploratory analysis used quantitative/semi-quantitative methods. We have grouped studies according to the metric used and reported the predictive accuracy of ¹⁸F-FDG PET for conversion to AD dementia at study level (Table 1). None of the included studies compared different validated analytical tools in the same research participants.

Eighteen studies used only computer aided visual read metrics (sc-SPM=8; Neurostat/3-D-SSP=6; SUVr/ROI=4); three [45, 50, 60] used only fully automated metrics; two used both approaches [24, 44]; two [46, 64] used principal component analysis on volume region of interest (VROI), and one [66] used a combination of SUVr images and t-sum. The highest values for both sensitivity and specificity (approximately ≥80%) were achieved in six

out of eight (2 'old' and 4 'new') that used the sc-SPM metrics [16, 39, 43, 47, 48, 55]. However, only one of these studies has a sample size above 50 participants and duration of follow-up of two years or longer [39]. For fully automated metrics, sensitivity values ranged from 61–79% while specificity values ranged from 29–91%. Three studies [44–45, 50], using these automated metrics, had a sample size above 50 participants and duration of follow-up two years or longer, but only one reported values for both sensitivity and specificity approximately ≥80% [50]. Supplementary Table 4 (Index test sections) shows how these studies differ in analytical and image analysis approaches.

ADNI studies

The results from ADNI studies are reported separately to illustrate the accuracy of ¹⁸F-FDG PET as a single test amongst studies that analysed data from the same cohort (Table 2). At the study level the sensitivity values ranged from 10–92% and the specificity values ranged from 55–97%. Values for both sensitivity and specificity were approximately ≥80% in only three studies [40, 51, 52], of which two used the sc-SPM metric.

Table 2
Accuracy figures of ¹⁸F-FDG PET for conversion from MCI to AD dementia at study levels in ADNI studies

Metric	Study ID	Participants (No. of ¹⁸ F-FDG PET positive*)	Age (y) All MCI participants or converters (non-converters)	Accuracy at study level			Duration of follow-up (months)
				Sensitivity (%)	Specificity (%)	No. with MCI converted (%)	
<i>Computer aided visual read</i>							
SPM							
¹⁸ F-FDG PET scan & covariates	Arbizu 2013	121 (65)	77 Range 65–88	92	62	36 (30)	85
	Dukart 2016	164 (48)	73 ± 7.8	90	84	29 (18)	135
¹⁸ F-FDG PET scan alone	Toussaint 2012	80 (44)	76 ± 4.1 Range 70–86	85	75	40 (50)	40
	Landau 2010	85 (51)	78 ± 7.5 (78 ± 7.4)	75	47	28 (33)	57
ROI/SUVr							
	Prestia 2013 ADNI cohort	57 (25)	75 ± 8 Range 55–89	46	58	24 (42)	33
	TOMC cohort	36 (13)	72 ± 8 Range 51–85	56	83	18 (50)	18
	Schmand 2012	89 (18)	74 ± 7.4 (74 ± 7.6) Baseline sample N = 175	24	82	38 (43)	51
	Trzepacz 2014	50 (3)	75 ± 6.6 (74 ± 8.4) Range 58–86	10	97	20 (40)	30
SVM							
Gaussian classification	Young 2013	143 (48)	69 ± 5.5 (68 ± 7.6)	55	77	47 (33)	96
		143 (65)		66	65		96

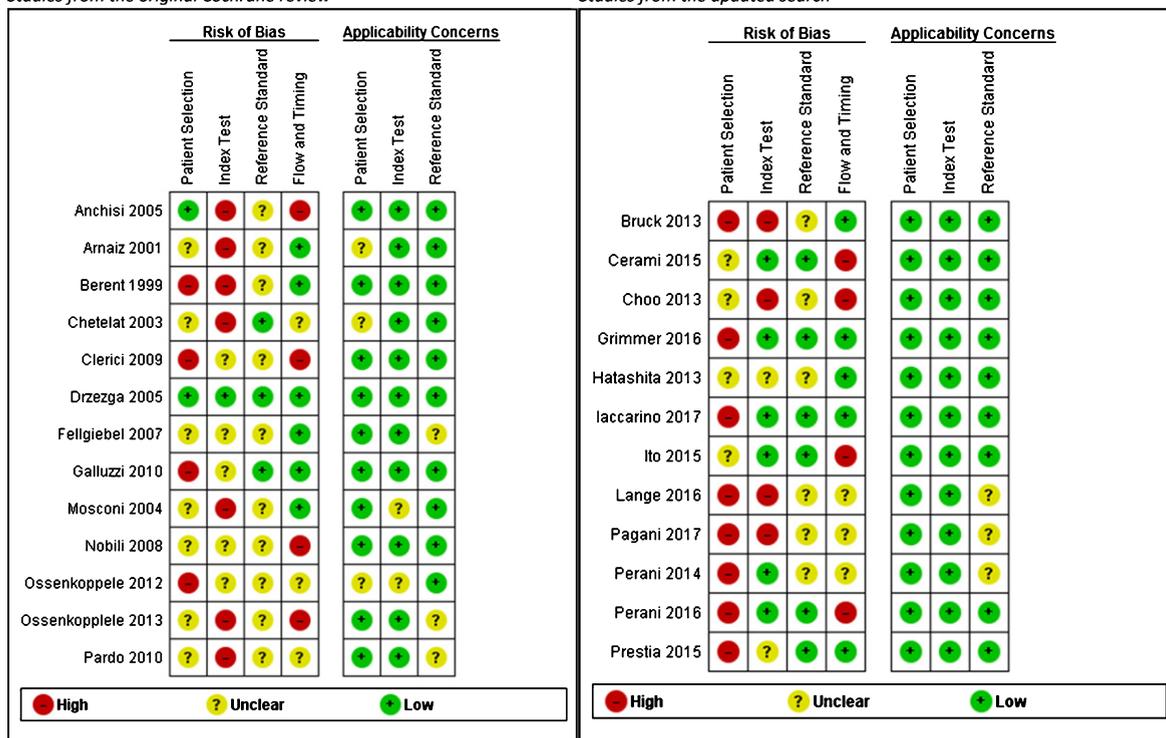
<i>Independent component analysis (ICA) (not computer aided visual read)</i>								
¹⁸ F-FDG PET scan alone	Schaffer 2012	97 (51)	75 ± 7.2 (75 ± 7.2)	88	76	43 (44)	54	48
¹⁸ F-FDG PET scan & covariates		97 (46)		81	83			
<i>Automated read</i>								
t-sum	Herholz 2011	94 (38)	75.0 ± 7.6	57	67	30 (32)	64	24
	Lange 2016	241 (100)	74 ± 6.5 (70 ± 7.2)	70	68	60 (25)	181	36
	Prestia 2013							
	ADNI cohort	57 (19)	75 ± 8 Range 55–89	33	67	24 (42)	33	mean 36 ± 12 range 12–48
	TOMC cohort	36 (19)	72 ± 8 Range 51–85	78	72	18 (50)	18	mean 26 ± 12 range 12–36
HCI	Gommar 2014	162 (59)	75 ± 7.3 Baseline sample N = 318	47	73	74 (46)	88	48 range 7.3–61.4
	Prestia 2013							
	ADNI cohort	57 (28)	75 ± 8 Range 55–89	63	61	24 (42)	33	mean 36 ± 12 range 12–48
	TOMC cohort	36 (21)	72 ± 8 Range 51–85	72	55	18 (50)	18	mean 26 ± 12 range 12–36

¹⁸F-FDG PET, fluorine-18-2-fluoro-2-deoxy-D-glucose positron emission tomography; MCI, mild cognitive impairment; AD, Alzheimer’s disease; SPM, statistical parametric map; SUVr, standardised uptake value ratio; ROI, region of interest; VROI, volume region of interest; HCI, hypometabolic convergence index; TOMC, Transitional Outpatient Memory Clinic; SVM, support vector machine. *¹⁸F-FDG PET positive relates to hypometabolism in brain region exceeding a certain threshold. **No overlap between participants in those two studies (Dr. Perani). * ADNI study, Alzheimer’s Disease Neuroimaging Initiative cohort.

Risk of bias and applicability concerns summary

Studies from the original Cochrane review

Studies from the updated search



Footnote: Assessment criteria using QUADAS tool; full details of quality assessment questions and individual study assessment in Supplementary Table 4

Fig. 4. Risk of bias and applicability concerns summary: review authors’ judgements about each domain for each included study.

Supplementary Table 5 shows that all studies applied a retrospective analysis of longitudinal data for ‘MCI-converters’ and ‘MCI-stable’. Sample sizes ranged from 50–241 participants. Duration of follow-up averaged two-three years; in four studies [40, 41, 49, 51] participants were followed-up for four years or longer. Most of the included studies used a validated voxel-based analysis method of ¹⁸F-FDG PET imaging [13].

The conclusions in ADNI studies differed. For instance, it is reported that for a single biomarker, when compared to other biomarkers, the best accuracy was obtained using ¹⁸F-FDG PET [40]. However, ¹¹C-PIB PET generated the best sensitivity, and ¹⁸F-FDG PET the lowest sensitivity in Trzepacz et al. [53]. PET score using ¹⁸F-FDG PET predicted clinical progression from MCI to AD dementia with a higher accuracy than Mini-Mental State Examination and Alzheimer’s Disease Assessment Scale cognitive subscale [49]. In contrast, cognitive markers were more robust predictors than biomarkers [41].

Quality assessment using QUADAS 2

QUADAS 2 scores for each domain are shown in Fig. 4 and further assessment details for studies included in the exploratory analyses are shown in Supplementary Table 4. Assessment of methodological quality was hampered by poor reporting and lack of details. Areas of particular concern for risk of bias were around *Patient Selection* and *Index Test* domains. Poor reporting about sampling procedure led mainly to unclear risk of bias or contributed to high risk of bias in the participant selection domain. For the Index test domain, eleven studies were of poor methodological quality and at high risk of bias due to lack of pre-specified threshold. Although the reference standard was regarded as adequate to correctly classify the target condition, poor reporting on blinding of dementia assessors resulted in unclear risk of bias for that domain in most of the included studies (*n* = 16).

In summary, most studies have major threats to their validity according to QUADAS 2. Only one

study [17] passed all the QUADAS 2 criteria; over half ($n=13$) failed on four or more of the seven available sets of criteria. The evidence base could thus be safely categorized as generally low quality.

A brief overview of the quality assessment is shown in the Legend section of Supplementary Table 4.

An overview and characteristics of studies included in the exploratory analysis are presented in Table 3.

DISCUSSION

This paper updates the previous ¹⁸F-FDG PET Cochrane review [19] to determine this biomarker's accuracy for prediction of conversion from MCI to AD dementia in light of recently published studies. Deploying the same methodological rigor, it yielded 36 studies overall. When all included studies were examined across all metrics, exploratory analysis for conversion of MCI to AD dementia ($n=24$) showed sensitivity values ranged from 25–100% while specificity values ranged from 24–100%. Highly variable accuracy for all metrics was also found across the eleven newly included studies, published in the last 5 years. Sensitivity values ranged from 56–100% and specificity values from 24–100%. The best values are reported for the sc-SPM metric in four new studies; sensitivity values ranged from 77–100% and specificity from 74–100%. Three [43, 47, 48] of these four studies used an optimized voxel-based procedure [69]. To be useful for dementia diagnosis, achieving both high sensitivity and specificity is critical [70].

The main sources of heterogeneity across studies are differences in study design, participants sampling, sample size and settings, variability in selected thresholds with respect to hypometabolism in different brain regions, variable length of follow-up, and analytical approaches (Table 3). The DTA study designs least likely to cause bias are prospective longitudinal studies with consecutive sampling and large sample size. Accuracy obtained from studies with retrospective design might be overestimated [71], and small samples may result in misleading estimates of test accuracy due to wide confidence intervals [72]. In this review, fourteen prospective cohort studies and ten that retrospectively analysed longitudinal data were included in the exploratory analysis for conversion to AD dementia. Only 5 studies recruited a consecutive sample. Most included studies had a sample size below 50, four between

50–120 participants, and only one study with over 200 participants; participants were recruited from a range of settings. A substantial number of included studies had limitations in methodological and reporting quality. Areas of particular bias concern were *Patient Selection* and *Index Test* domains; only one study was graded low risk of bias for all criteria of QUADAS 2 [17] and reported high values for sensitivity (92%) and specificity (89%) using Neurostat/3D-SSP metric. However, this study has a small sample size ($n=30$); therefore, the accuracy achieved might be overestimated.

The accuracy of biomarkers also heavily depends on the length of the follow-up period. MCI participants with a positive (abnormal) ¹⁸F-FDG PET scan are classified as 'false positive' if they do not convert to AD dementia at a predefined follow-up period. It is possible the specificity values in some included studies with short follow-up are falsely low because some MCI participants might only convert to dementia at a later follow-up period and would thus be classified 'true positive'. The same logic applies to sensitivity values with the 'misclassification' of participants as true negative. But what is the ideal follow-up to assess diagnostic accuracy? Duration of follow-up was 24–36 months in most included studies and one study [46] followed-up participants for five years. Three studies with a relatively short follow-up of 18 months or less [16, 17, 63] reported specificity values of 100%, 89%, and 97%, respectively. One plausible explanation is that MCI study participants had more severe cognitive impairment at baseline, and consequently converted to AD dementia during the short observation period. Despite the fact that a majority of included studies (33/36) used Petersen's standardized clinical criteria for MCI diagnosis [4, 27], reported information about participants' cognitive impairment severity at baseline was insufficient to explore this issue. A better understanding of the role of biomarkers in prediction of conversion from MCI stage to AD requires both short and long-term periods of observation [25] and characterization of the severity of cognitive decline at baseline.

Increasing age of MCI participants is considered the strongest risk factor for progression to AD dementia [73, 74]. Participants' age for studies included in the exploratory analysis ranged from 50–83 years. The influence of age groups on the predictive accuracy of ¹⁸F-FDG PET was assessed only in a few included studies. One study reported that age as covariate in statistical analysis did not improve the performance of ¹⁸F-FDG PET [45]; another found

Table 3
Summary of the main findings

Research question: What is the diagnostic accuracy of ¹⁸ F-FDG PET biomarker for detecting AD pathology <i>in vivo</i> and predicting progression to AD dementia in people with MCI	
Studies included in an exploratory analysis for conversion from MCI to AD dementia (<i>n</i> = 24)	
Study design	Prospective study design (<i>n</i> = 14): longitudinal cohort study (<i>n</i> = 8) and nested case-control study (<i>n</i> = 6)
Participant population/AGE	Retrospective study design (<i>n</i> = 10): retrospective analysis of longitudinal data Participants diagnosed with MCI at baseline Age ranged from 50 to 83 years
MCI criteria	Petersen criteria (<i>n</i> = 21) Global Deterioration Scale (GDS) (<i>n</i> = 1) (old study) [57] Isolated memory impairment (IMI) criteria (<i>n</i> = 1) (old study) [58] Neuropsychological set battery for aMCI (<i>n</i> = 1) (new study) [45] Notes: a stage of MCI (earlier/advanced) was not generally reported in primary studies
Sampling procedure	Consecutive (<i>n</i> = 4) Not consecutive (12) Unclear (<i>n</i> = 8)
Sample size	A number included in the analysis in primary studies ranged from 12 to 241 participants Less than 30 participants (range: 12–28) (<i>n</i> = 12) 30–50 participants (range: 30–48) (<i>n</i> = 6) More than 50 participants (range: 68–241) (<i>n</i> = 6)
Settings	Most of the studies (<i>n</i> = 15) recruited participants from Outpatient University Departments (e.g., memory clinics, cognitive disorders clinics, centers for research and treatment of cognitive dysfunctions, etc.) Multicenter (<i>n</i> = 2) Not reported (<i>n</i> = 7)
Index test	¹⁸ F-FDG PET
Threshold	Included studies used a range of thresholds regarding hypometabolism in different brain regions. Most of them were based on computer aided visual read (<i>n</i> = 17), while few of them were based only on fully automated measures of glucose hypometabolism (<i>n</i> = 3). Two studies applied both approaches. Two studies used principal component analysis on VROI, and one study combined SUVr images and t-sum.
Analytical approach (metric)	Neurostat/3D-SSP (<i>n</i> = 4) SPM (<i>n</i> = 8) SUVr/ROI (<i>n</i> = 5) Fully automated read (<i>n</i> = 4) Neurostat/3D-SSP and fully automated read (<i>n</i> = 2) Combined 'SUVr images' and 't-sum' (<i>n</i> = 1) VROI (<i>n</i> = 2)
Threshold pre-specified at baseline	Yes (<i>n</i> = 11) No/Unclear (<i>n</i> = 13)
¹⁸ F-FDG hypometabolism regions	Metabolism was studied in a range of brain areas. In two studies it is not specified which exact areas were investigated. Regarding the rest twenty-two studies, all of them involved the temporal/parietal cortex. Seventeen studies involved also the posterior cingulate cortex, while twelve of them involved frontal regions in their evaluations. These were the most commonly investigated brain regions in the included studies.
Reference standard	For AD dementia: NINCDS-ADRDA (<i>n</i> = 20); CDR = 1 (<i>n</i> = 1); 'non-specified clinical dementia rating' (<i>n</i> = 1); 'not reported' (<i>n</i> = 2) McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; NINDS-AIREN criteria for vascular dementia
Target condition	Conversion from MCI to AD dementia
Duration of follow-up	Less than 24 months (<i>n</i> = 3) On average, mainly between 24–36 months (<i>n</i> = 20); some participants were followed slightly longer 5 years or longer (<i>n</i> = 1)

(Continued)

Table 3
(Continued)

Included studies	Prospectively and retrospectively defined cohorts or nested case-control samples of MCI participants diagnosed mainly by Petersen criteria (21/24). Twenty-four studies are included in the exploratory for conversion to AD dementia (all metrics; all studies, $n = 1132$ participants; old studies, $n = 327$ and new studies, $n = 805$).
Quality concerns	QUADAS 2 scoring was challenging due to insufficient details. Poor reporting about sampling procedure led mainly to unclear risk of bias or contributed to high risk of bias in the participant selection domain. Although the reference standard was regarded as adequate to correctly classify the target condition, poor reporting on blinding of dementia assessors determined unclear risk of bias in the reference domain in most of the included studies ($n = 16$). According to the assessment of Index test domain, eleven studies were of poor methodological quality and were considered to be at high risk of bias due to lack of pre-specified threshold. Over a half ($n = 13$) failed on four or more of the seven available sets of criteria. Only one study [17] passed all sets of QUADAS 2 criteria. The evidence base could thus be safely categorized as generally low quality.
Heterogeneity	Studies included in the exploratory analysis for conversion to AD differ regarding i) some aspects of study design (e.g., prospective/retrospective design; consecutive/non-consecutive sampling; lack of information about severity of cognitive impairment in MCI); ii) duration of follow-up; iii) absence of common thresholds; a range of different thresholds used; iv) methodological differences and heterogeneity in the conduct and interpretation of the test regarding analytical approach used, brain areas investigated, scaling procedures, etc. These differences resulted in considerable heterogeneity between included studies. A meta-analysis was not performed, and consequently, a formal investigation of heterogeneity was not feasible.
Accuracy at study level	Overall, when all studies were examined across all metrics, the estimates of sensitivity and specificity for conversion to AD dementia vary considerably between included studies, and are likely to be attributable to covariates mentioned above. In the thirteen studies from previous review, the sensitivity values ranged from 25% to 100% while the specificity values ranged from 15% to 100%. In the eleven new identified studies, the values ranged from 56% to 100% and from 24% to 100% for sensitivity and specificity respectively. <i>sc-SPM metric</i> Sensitivity = 38% to 100%; Specificity = 15% to 100% (4 old and 4 new studies) Sensitivity = 77% to 100%; Specificity = 74% to 100% (4 new studies) <i>Neurostat/3D-SSP metric</i> Sensitivity = 25% to 100%; Specificity = 40% to 89% (4 old and 2 new studies) Sensitivity = 56% to 100%; Specificity = 40% to 74% (2 new studies) <i>t-sum/HCI metric</i> Sensitivity = 79%; Specificity = 29% (1 old study) Sensitivity = 61% to 79%; Specificity = from 37% to 91% (4 new studies) <i>Notes:</i> The most consistently high sensitivity and specificity (approximately $\geq 80\%$) were reported in the six sc-SPM studies (2 old and 4 new).

¹⁸F-FDG PET, fluorine 18-2-fluoro-2-deoxy-D-glucose positron emission tomography; MCI, mild cognitive impairment; VROI, volume region of interest; SUVr, standardized uptake value ratio; 3D-SSP, three-dimensional stereotactic surface projection; SPM, statistical parametric map; NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; CDR, clinical dementia rating; NINDS-AIREN, National Institute of Neurological Disorders and Stroke and the Association Internationale pour la Recherche et l'Enseignement en Neurosciences; AD, Alzheimer's disease; QUADAS, Quality Assessment of Diagnostic Accuracy Studies; HCI, hypometabolic convergence index; ADNI, Alzheimer's Disease Neuroimaging Initiative cohort.

that in participants under 75 years this biomarker was only a marginally significant predictor and in participants over 75 did not predict conversion [68].

Only ten out of 24 studies defined a specific threshold for ¹⁸F-FDG PET biomarker at baseline and then prospectively assessed its predictive accuracy in identifying those participants with MCI who would convert to AD dementia at follow-up. The remaining studies used optimized thresholds, so reported

sensitivity and/or specificity values might be overly optimistic. Furthermore, the accuracy of imaging biomarkers is highly influenced by the analytical approach used. It is still a matter of debate which ¹⁸F-FDG PET metrics are most accurate [25, 44]. Our review confirms previous evidence from a published meta-analysis [75] that sc-SPM metric shows the best prognostic accuracy when compared to other computer added visual read or fully automated metrics.

However, the findings should be interpreted with caution because: 1) only one study [40] had a sample size above 50 and follow-up of 2 years or longer as recommended [30]; 2) the accuracy of sc-SPM and other metrics was not assessed simultaneously in the same study participants; 3) four out of eight studies retrospectively analysed longitudinal data of MCI-converters and MCI-stable participants. Retrospective analysis has greater potential for bias than prospective longitudinal studies. A recent study [76], with 80 participants recruited retrospectively, concluded that an optimized voxel-based procedure sc-SPM metric has a relevant role in predicting progression to different dementias and in the exclusion of progression in prodromal MCI phase. These important, promising results for ¹⁸F-FDG PET-SPM need to be confirmed in future prospective longitudinal studies.

A number of studies [20, 30] concluded that a large body of data supports the accuracy of ¹⁸F-FDG PET to detect AD at MCI stage, showing high predictive value for conversion and providing highly relevant prognostic information for routine clinical use [30]. Based on a June 2015 search, Garibotto [30] reported that sensitivity ranged from 57–85% and specificity from 67–91% across five studies [44, 51, 52, 54, 61] with over 50 participants and a minimum two-year follow-up; four of these five studies included overlapping ADNI participants. Conversely, our review found that the accuracy of ¹⁸F-FDG PET as a single test among 12 studies from the same ADNI cohort varied immensely. Sensitivity values ranged from 10–92% and specificity from 55–97% across several metrics; values for both sensitivity and specificity were >80% in only three studies. When applying the same criteria regarding sample size and duration of follow-up, our review identified five additional studies [39, 42, 45, 46, 50] which reported sensitivity ranging from 63–93% and specificity from 24–93%; values for both sensitivity and specificity were $\geq 80\%$ in only three studies. Our comprehensive, up-to-date review supports Frisoni's statement [77] that "the informative value of biomarker cannot be used with full reliability in clinical practice" and further evidence of the clinical validity and utility of ¹⁸F-FDG PET in people with MCI is needed.

This review has systematically and comprehensively examined the up-to-date evidence for ¹⁸F-FDG PET as a single test for prediction of conversion to AD dementia in people with MCI, using Cochrane methods designed to minimize risk of bias for assessment and review of diagnostic test accuracy.

Overall, sensitivity, and specificity vary widely between studies, and this variability was still present in studies published in the last 5 years. However, there is some evidence that sc-SPM is the metric showing the best prognostic accuracy when compared to other computer aided visual read or fully automated metrics. The highest values for both sensitivity and specificity (approximately $\geq 80\%$) were achieved in 6 out of 8 studies (2 older and 4 new) that used the sc-SPM metrics. These findings should be interpreted with caution as only one study had a sample size above $n = 50$ and follow-up of 2 years or longer, and 4 out of 8 studies conducted retrospective analyses which are at greater risk of bias.

There are still methodological limitations in the available evidence and a lack of well-designed studies that meet best practice criteria for diagnostic test accuracy studies. Further work needs to be completed before ¹⁸F-FDG-PET as a single test can be widely recommended as a routine diagnostic test for conversion from MCI to AD dementia in clinical practice.

The promising results need further testing and confirmation in robust, methodologically sound prospective longitudinal cohort studies with long (≥ 5 years) follow-up, with defined baseline threshold(s) and larger consecutive samples stratified by age groups and other covariates. Various aspects of more efficient metrics, such as sc-SPM, need to be harmonized [30], optimized, standardized, agreed amongst experts and further tested. The predictive accuracy and the incremental diagnostic values of different metrics should also be assessed simultaneously in the same MCI participants. The use of dementia-specific guidance such as those proposed by STARDdem [78] may improve reporting quality in further DTA studies.

ACKNOWLEDGMENTS

We thank Isla Kuhn, Medical Librarian, for help with the literature searches. We are also grateful to Andy Cowan, Communications & Project Administrator, for help with editing and the submission process. This paper presents independent research funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East of England (Grant number: RG74482). The views expressed are those of the authors and not necessarily those of the National Health System, the NIHR, the Department of Health and Social Care.

Authors' disclosures available online (<https://www.j-alz.com/manuscript-disclosures/17-1125r1>).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/JAD-171125>.

REFERENCES

- [1] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-944.
- [2] Prince M, Bryce R, Ferri C (2011) World Alzheimer Report 2011: The benefits of early diagnosis and intervention. Alzheimer Disease International <https://www.alz.co.uk/research/WorldAlzheimerReport2011.pdf>, accessed on March 28, 2017.
- [3] Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G (2016) Timely diagnosis for Alzheimer's disease: A literature review on benefits and challenges. *J Alzheimers Dis* **49**, 617-631.
- [4] Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med* **256**, 183-194.
- [5] Bruscoli M, Lovestone S (2004) Is MCI really just early dementia? A systematic review of conversion studies. *Int Psychogeriatr* **16**, 129-140.
- [6] Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, Herukka SK, van der Flier WM, Blankenstein MA, Ewers M, Rich K, Kaiser E, Verbeek M, Tsolaki M, Mulugeta E, Rosén E, Aarsland D, Visser PJ, Schröder J, Marcusson J, de Leon M, Hampel H, Scheltens P, Pirttilä T, Wallin A, Jönhagen ME, Minthon L, Winblad B, Blennow K (2009) CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* **302**, 385-393.
- [7] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol* **56**, 303-308.
- [8] Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, Smith GE, Jack CR (2009) Mild cognitive impairment: Ten years later. *Arch Neurol* **66**, 1447-1455.
- [9] Vos SJB, Xiong C, Visser PJ, Jasielec MS, Hassenstab J, Grant EA, Cairns NJ, Morris JC, Holtzman DM, Fagan AM (2013) Preclinical Alzheimer's disease and its outcome: A longitudinal cohort study. *Lancet Neurol* **12**, 957-965.
- [10] Minoshima S, Giordani B, Berent S, Frey KA, Foster NL, Kuhl DE (1997) Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neurol* **42**, 85-94.
- [11] Nitirini R, Buchpiguel C, Caramelli P, Bahia VS, Mathias SC, Nascimento CM, Degenszajn J, Caixeta L (2000) SPECT in Alzheimer's disease: Features associated with bilateral parietotemporal hypoperfusion. *Acta Neurol Scand* **101**, 172-176.
- [12] Mosconi L (2013) Glucose metabolism in normal aging and Alzheimer's disease: Methodological and physiological considerations for PET studies. *Clin Transl Imaging* **1**, 217-233.
- [13] Perani D, Iaccarino L, Bettinardi V (2014) The need for "objective measurements" in FDG and amyloid PET neuroimaging. *Clin Transl Imaging* **2**, 331-342.
- [14] Mosconi L (2009) FDG PET in the evaluation of mild cognitive impairment and early dementia. In *PET in the Evaluation of Alzheimer's Disease and Related Disorders*, Silverman D, ed. Springer, New York, pp. 49-65.
- [15] Silverman DHS, Small GW, Chang CY, Lu CS, de Aburto MA, Chen W, Czernin J, Rapoport SI, Pietrini P, Alexander GE, Schapiro MB, Jagust WJ, Hoffman JM, Welsh-Bohmer KA, Alavi A, Clark CM, Salmon E, de Leon MJ, Mielke R, Cummings JL, Kowell AP, Gambhir SS, Hoh CK, Phelps ME (2001) Positron emission tomography in evaluation of dementia: Regional brain metabolism and long-term outcome. *JAMA* **286**, 2120-2127.
- [16] Chételat G, Desgranges B, De la Sayette V, Viadre F, Eustache F, Baron JC (2003) Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? *Neurology* **60**, 1374-1377.
- [17] Drzezga A, Grimmer T, Rimenschneider M, Lautenschlager N, Siebner H, Alexopoulos P, Minoshima S, Schwaiger M, Kurz A (2005) Prediction of individual clinical outcome in MCI by means of genetic assessment and (18)F-FDG PET. *J Nucl Med* **46**, 1625-1632.
- [18] Patterson JC, Lilien DL, Takalkar A, Pinkston JB (2011) Early detection of brain pathology suggestive of early AD using objective evaluation of FDG-PET scans. *Int J Alzheimers Dis* **2011**, 946590.
- [19] Smailagic N, Vacante M, Hyde C, Martin S, Ukoumunne O, Sachpekidis C (2015) ¹⁸F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* **1**, CD010632.
- [20] Morbelli S, Garibotto V, Van De Giessen E, Arbizu J, Chételat G, Drezgza A, Hesse S, Lammertsma AA, Law I, Payoux P, Pagani M (2015) A Cochrane review on brain ¹⁸F-FDG PET in dementia: Limitations and future perspectives. *Eur J Nucl Med Mol Imaging* **42**, 1487-1491.
- [21] Beynon R, Leeftang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, Glanville JM (2013) Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database Syst Rev*, MR000022.
- [22] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps C (2011) The diagnosis of mild cognitive impairment 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* **7**, 270-279.
- [23] Cabral C, Morgado PM, Costa DC, Silveira M; Alzheimer's Disease Neuroimaging Initiative (2015) Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. *Comput Biol Med* **58**, 101-109.
- [24] Grimmer T, Wutz C, Alexopoulos P, Drzezga A, Förster S, Förstl H, Goldhardt O, Ortner M, Sorg C, Kurz A (2016) Visual versus fully automated analyses of ¹⁸F-FDG and amyloid PET for prediction of dementia due to Alzheimer's disease in mild cognitive impairment. *J Nucl Med* **57**, 204-207.

- [25] Morbelli S, Brugnolo A, Bossert I, Buschiazzo A, Frisoni GB, Galluzzi S, Van Berckel BN, Ossenkoppele R, Pernecky R, Drzezga A, Didic M, Guedj E, Sambuceti G, Bottoni G, Arnaldi D, Picco S, De Carli F, Pagani M, Nobili (2015) Visual versus semi-quantitative analysis of ¹⁸F-FDG PET in amnesic MCI: An European Alzheimer's disease Consortium (EADC) project. *J Alzheimers Dis* **44**, 815-826.
- [26] Zhang S, Han D, Tan X, Feng J, Guo Y, Ding Y (2012) Diagnostic accuracy of ¹⁸F-FDG PET for prediction of short-term conversion to Alzheimer's disease in subjects with mild cognitive impairment. *Int J Clin Pract* **66**, 185-198.
- [27] Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, Nordberg A, Bäckman L, Albert M, Almkvist O, Arai H, Basun H, Blennow K, de Leon M, DeCarli C, Erkinjuntti T, Giacobini E, Graff C, Hardy J, Jack C, Jorm A, Ritchie K, van Duijn C, Visser P, Petersen RC (2004) Mild cognitive impairment—beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* **256**, 240-246.
- [28] Matthews FE, Stephan B, McKeith IG, Bond J, Brayne C (2008) Two-year progression from mild cognitive impairment to dementia: To what extent do different definitions agree? *J Am Ger Soc* **56**, 1424-1433.
- [29] Mason SE, McShane R, Ritchie CW (2010) Diagnostic tests for Alzheimer's disease: Rationale, methodology and challenges. *Int J Alzheimers Dis* **2010**, 972685.
- [30] Garibotto V, Herholz K, Boccardi M, Picco A, Varrone A, Nordberg A, Nobili F, Ratib O, Geneva Task Force for the Roadmap of Alzheimer's Biomarkers (2017) Clinical validity of brain fluorodeoxyglucose positron emission tomography as biomarker for Alzheimer's disease in the context of a structured 5-phase development frame work. *Neurobiol Aging* **52**, 183-195.
- [31] World Health Organization (2010) Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10). <http://apps.who.int/classifications/icd10/browse/2010/en>, accessed on December 16, 2012.
- [32] American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. 4th edition, American Psychiatric Association, Washington DC.
- [33] Review Manager (RevMan) Version 5.3 [Computer program] (2014) Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.
- [34] Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* **58**, 982-990.
- [35] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM (2011) QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* **155**, 529-536.
- [36] Arbizu J, Prieto E, Martínez-Lage P, Martí-Climent JM, García-Granero M, Lamet I, Pastor P, Riverol M, Gómez-Isla MT, Peñuelas I, Richter JA, Weiner MW, for the Alzheimer's Disease Neuroimaging Initiative (2013) Automated analysis of FDG PET as a tool for single-subject probabilistic prediction and detection of Alzheimer's disease dementia. *Eur J Nucl Med Mol Imaging* **40**, 1394-1405.
- [37] Bruck A, Virta JR, Koivunen J, Scheinin NM, Helenius H, Nägren K, Helin S, Parkkola R, Viitanen M, Rinne JO (2013) [¹¹C], [¹⁸F]FDG and MR imaging in patients with mild cognitive impairment. *Eur J Nucl Med Mol Imaging* **40**, 1567-1572.
- [38] Cerami C, Della Rosa PA, Magnani G, Santangelo R, Marccone A, Cappa SF, Perani D (2015) Brain metabolic maps in mild cognitive impairment predicts heterogeneity of progression to dementia. *Neuroimage Clin* **7**, 187-194.
- [39] Choo ILH, Ni R, Scholl M, Wall A, Almkvist O, Nordberg A (2013) Combination of ¹⁸F-FDG PET and cerebrospinal fluid biomarkers as a better predictor of the progression to Alzheimer's disease in mild cognitive impairment. *J Alzheimers Dis* **33**, 929-939.
- [40] Dukart J, Sambataro F, Bertolino A (2016) Accurate prediction of conversion to Alzheimer's disease dementia using imaging, genetic and neuropsychological biomarkers. *J Alzheimers Dis* **49**, 1143-1159.
- [41] Gommer JJ, Conejero-Goldberg C, Davies P, Goldberg TE (2014) Extension and refinement of the predictive value of different classes of markers in ADNI: Four-year follow-up data. *Alzheimers Dement* **10**, 704-712.
- [42] Hatashita S, Yamasaki H (2013) Diagnosed mild cognitive impairment due to Alzheimer's disease with PET biomarkers of Beta amyloid and neuronal dysfunction. *Neurology* **8**, e66877602.
- [43] Iaccarino L, Chiotis K, Alongi P, Almkvist O, Wall A, Cerami C, Bettinardi V, Gianolli L, Nordberg A, Perani D (2017) A cross-validation of FDG- and amyloid-PET biomarkers in mild cognitive impairment for the risk prediction to dementia due to Alzheimer's disease in clinical setting. *J Alzheimers Dis* **59**, 603-614.
- [44] Ito K, Fukuyama H, Senda M, Ishii K, Maeda K, Yamamoto Y, Ouchi Y, Ishii K, Okumura A, Fujiwara K, Kato T, Arahata Y, Washimik Y, Mitsuyamal Y, Megurom K, Ikedan M, SEAD-J Study Group (2015) Prediction of outcomes in mild cognitive impairment by using ¹⁸F-FDG PET: A multicentre study. *J Alzheimers Dis* **45**, 543-552.
- [45] Lange C, Suppa P, Frings L, Brenner W, Spies L, Buchert R, for the Alzheimer's Disease Neuroimaging Initiative (2016) Optimization of statistical single subject analysis of brain FDG-PET for the prognosis of mild cognitive impairment-to-Alzheimer's disease conversion. *J Alzheimers Dis* **49**, 945-959.
- [46] Pagani M, Nobili F, Morbelli S, Arnaldi D, Giuliani A, Öberg J, Girtler N, Brugnolo A, Picco A, Bauckneht M, Piva R, Chincari A, Sambuceti G, Jonsson C, De Carli F (2017) Early identification of MCI converting to AD: FDG PET study. *Eur J Nucl Med Mol Imaging* **44**, 2042-2052.
- [47] Perani D, Della Rosa PA, Cerami C, Gallivanone F, Fallanca F, Vanoli EG, Panzacchi A, Nobili F, Pappata S, Marccone A, Garibotto V, Castiglioni I, Magnani G, Cappa SF, Gianolli L; EADC-PET Consortium (2014) Validation of an optimized SPM procedure for the FDG-PET in dementia diagnosis in clinical setting. *Neuroimage Clin* **6**, 445-454.
- [48] Perani D, Cerami C, Caminiti SP, Santangelo R, Coppi E, Ferrari L, Pinto P, Passerini G, Falini A, Iannaccone S, Cappa SF, Comi G, Gianolli L, Magnani G (2016) Cross-validation of biomarkers for the early differential diagnosis and prognosis of dementia in a clinical setting. *Eur J Nucl Med Mol Imaging* **43**, 499-508.
- [49] Prestia A, Caroli A, Herholz K, Reiman E, Chen K, Jagust WJ, Frisoni GB, Translational Outpatient Memory Clinic Working Group, Alzheimer's Disease Neuroimaging

- Initiative (2013) Diagnostic accuracy of markers for prodromal Alzheimer's disease in independent clinical series. *Alzheimers Dement* **9**, 677-686.
- [50] Prestia A, Caroli A, Wade SK, van der Flier WM, Ossenkoppele R, Van Berckel B, Barkhof F, Teunissen CE, Wall A, Carter SF, Schöll M, Choo IH, Nordberg A, Scheltens P, Frisoni GB (2015) Prediction of AD dementia by biomarkers following the NIA-AA and IWG diagnostic criteria in MCI patients from three European memory clinics. *Alzheimers Dement* **11**, 1191-1201.
- [51] Shaffer JL, Petrella JR, Sheldon FC, Choudhury KR, Calhoun VD, Coleman RE, Doraiswamy PM (2013) Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology* **266**, 583-591.
- [52] Toussaint P-J, Perlberg V, Bellec P, Desarnaud S, Lacomblez L, Doyon J, Habert MO, Benali H (2012) Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. *Neuroimage* **63**, 936-946.
- [53] Trzepacz PT, Yu P, Sun J, Schuh K, Case M, Witte MM, Hochstetler H (2014) Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. *Neurobiol Aging* **35**, 143-151.
- [54] Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S (2013) Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin* **2**, 735-745.
- [55] Anchisi D, Borroni B, Franceschi M, Kerrouche N, Kalbe E, Beuthien-Beumann B, Cappa S, Lenz O, Ludecke S, Marcone A, Mielke R, Ortelli P, Padovani A, Pelati O, Pupi A, Scarpini E, Weisenbach S, Herholz K, Salmon E, Holthoff V, Sorbi S, Fazio F, Perani D (2005) Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease. *Arch Neurol* **62**, 1728-1733.
- [56] Arnáiz E, Jelic V, Almkvist O, Wahlund LO, Winblad B, Valind S, Nordberg A (2001) Impaired cerebral glucose metabolism and cognitive functioning predict deterioration in mild cognitive impairment. *Neuroreport* **12**, 851-855.
- [57] Berent S, Giordani B, Foster N, Minoshima S, Lajiness-O'Neill R, Koeppe R, Kuhl D (1999) Neuropsychological function and cerebral glucose utilization in isolated memory impairment and Alzheimer's disease. *J Psychiatr Res* **33**, 7-16.
- [58] Clerici F, Del Sole A, Chiti A, Maggiore L, Lecchi M, Pomati S, Mosconi L, Lucignani G, Mariani C (2009) Differences in hippocampal metabolism between amnesic and non-amnesic MCI subjects: Automated FDG-PET image analysis. *Q J Nucl Med Mol Imaging* **53**, 646-657.
- [59] Fellgiebel A, Scheurich A, Bartenstein P, Muller MJ (2007) FDG-PET and CSF phospho-tau for prediction of cognitive decline in mild cognitive impairment. *Psychiatry Res* **155**, 167-171.
- [60] Galluzzi S, Geroldi C, Ghidoni R, Paghera B, Amicucci G, Bonetti M, Zanetti O, Cotelli M, Gennarelli M, Frisoni GB, Translational Outpatient Memory Clinic Working Group (2010) The new Alzheimer's criteria in a naturalistic series of patients with mild cognitive impairment. *J Neurol* **257**, 2004-2014.
- [61] Herholz K, Westwood S, Haense C, Dunn G (2011) Evaluation of a calibrated (18)F-FDG PET score as a biomarker for progression in Alzheimer disease and mild cognitive impairment. *J Nucl Med* **52**, 1218-1226.
- [62] Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, Petersen RC, Shaw LM, Trojanowski JQ, Jack CR, Weiner MW, Jaquet WJ (2010) Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* **75**, 230-238.
- [63] Mosconi L, Perani D, Sorbi S, Herholz K, Nacmias B, Holthoff V, Salmon E, Baron JC, De Cristofaro MT, Padovani A, Borroni B, Franceschi M, Bracco L, Pupi A (2004) MCI conversion to dementia and APOE genotype: A prediction study with FDG-PET. *Neurology* **63**, 2332-2340.
- [64] Nobili F, Salmaso D, Morbelli S, Girtler N, Piccardo A, Brugnolo A, Dessi B, Larsson SA, Rodriguez G, Pagani M (2008) Principal component analysis of FDG PET in amnesic MCI. *Eur J Nucl Med Mol Imaging* **35**, 2191-2202.
- [65] Ossenkoppele R, Tolboom N, Foster-Dingley JC, Adriaanse SF, Boellaard R, Yaquib M, Windhorst AD, Barkhof F, Lammertsma AA, Scheltens P, van der Flier WM, van Berckel BN (2012) Longitudinal imaging of Alzheimer pathology using [11C]PIB, [18F]FDNP and [18F]FDG PET. *Eur J Nucl Med Mol Imaging* **39**, 990-1000.
- [66] Ossenkoppele R, Prins N, Pijnenburg YAL, 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 (2013) Impact of molecular imaging on the diagnostic process in a memory clinic. *Alzheimers Dement* **9**, 414-421.
- [67] Pardo JV, Lee JT, Kuskowski MA, Munch KR, Carlis JV, Sheikh SA, Surerus C, Lewis SM, McCarten JR, Fink H, McPherson S, Shah HH, Rottunda S, Dysken MW (2010) Fluorodeoxyglucose positron emission tomography of mild cognitive impairment with clinical follow-up at 3 years. *Alzheimers Dement* **6**, 326-333.
- [68] Schmand B, Eikelenboom P, van Gool WA; Alzheimer's Disease Neuroimaging Initiative (2012) Value of diagnostic tests to predict conversion to Alzheimer's disease in young and old patients with amnesic mild cognitive impairment. *J Alzheimers Dis* **29**, 641-648.
- [69] Della Rosa PA, Cerami C, Gallivanone F, Prestia A, Caroli A, Castiglioni I, Gilardi MC, Frisoni G, Friston K, Ashburner J, Perani D; the EADC-PET Consortium (2014) A standardized [18F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. *Neuroinformatics* **12**, 575-593.
- [70] Humpel C (2011) Identifying and validating biomarkers for Alzheimer's disease. *Trends Biotechnol* **29**, 26-32.
- [71] Whiting PF, Ruthjes AWS, Westwood ME, Mallett S, QUADAS-2 Steering Group (2013) A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* **66**, 1093-1104.
- [72] Bochmann F, Johnson Z, Azuara-Blanco A (2007) Sample size in studies on diagnostic accuracy in ophthalmology: A literature survey. *Br J Ophthalmol* **91**, 898-900.
- [73] Visser PJ, Kester A, Jolles J, Verhey F (2006) Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology* **67**, 1201-1207.
- [74] Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C (2009) Age, neuropathology, and dementia. *N Engl J Med* **360**, 2302-2309.
- [75] Frisoni GB, Bocchetta M, Chetelat G, Rabinovici GD, De Leon MJ, Kaye J, Reiman EM, Scheltens P, Barkhof F, Black SE, Brooks DJ, Carrillo MC, Fox NC, Herholz K, Nordberg

- A, Jack CR Jr, Jagust WJ, Johnson KA, Rowe CC, Sperling RA, Thies W, Wahlund LO, Weiner MW, Pasqualetti P, Decarli C (2013) Imaging markers for Alzheimer disease. Which vs how. *Neurology* **81**, 487-500.
- [76] Caminiti SP, Ballarini T, Sala A, Cerami C, Presotto L, Santangelo R, Fallanca F, Vanoli EG, Gianolli L, Iannaccone S, Magnani G, Perani D; BIOMARKAPD Project (2018) FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort. *Neuroimage Clin* **18**, 167-177.
- [77] Frisoni GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, Démonet J-F, Garibotto V, Giannakopoulos P, Gietl A, Hansson O, Herholz K, Jack CR Jr, Nobili F, Nordberg A, Snyder HM, Kate MT, Varrone A, Albanese E, Becker S, Bossuyt P, Carrillo MC, Cerami C, Dubois B, Gallo V, Giacobini E, Gold G, Hurst S, Lönneborg A, Lovblad K-O, Mattsson N, Molinuevo J-L, Monsch AU, Mosimann U, Padovani A, Picco A, Porteri C, Ratib O, Saint-Aubert L, Scerri C, Scheltens P, Schott JM, Sonni I, Teipel S, Vineis P, Visser PJ, Yasui Y, Winblad B (2017) Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *Lancet Neurol* **16**, 661-676.
- [78] Noel-Storr AH, McCleery JM, Richard E, Ritchie CW, Flicker L, Cullum SJ, Davis D, Quinn TJ, Hyde C, Rutjes AW, Smailagic N, Marcus S, Black S, Blennow K, Brayne C, Fiorivanti M, Johnson JK, Köpke S, Schneider LS, Simmons A, Mattsson N, Zetterberg H, Bossuyt PM, Wilcock G, McShane R (2014) Reporting standards for studies of diagnostic test accuracy in dementia. *Neurology* **83**, 364-373.