Supplementary Table 4

**QUADAS 2 assessments criteria: Anchoring statements for quality assessment of 18F-FDG PET DTA studies**

|  |  |  |
| --- | --- | --- |
| **Question** | **Response and weighting** | **Explanation** |
| **Patient Selection** | | |
| Was the sampling method appropriate? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *Where sampling is used, the designs least likely to cause bias are consecutive sampling or random sampling. Sampling that is based on volunteers or selecting participants from a clinic or research resource is prone to bias.* |
| Was a case-control or similar design avoided? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *Designs similar to case-control that may introduce bias are those designs where the study team deliberately increase or decrease the proportion of participants with the target condition, which may not be representative. Some case-control methods may already be excluded if they mix participants from various settings.* |
| Are exclusion criteria described and appropriate? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *Study will be automatically graded unclear if exclusions are not detailed (pending contact with study authors). Where exclusions are detailed, the study will be graded as 'low risk' if exclusions are felt to be appropriate by the review authors. Certain exclusions common to many studies of dementia are: medical instability; terminal disease; alcohol/substance misuse; concomitant psychiatric diagnosis; other neurodegenerative condition. Exclusions are not felt to be appropriate if ‘difficult to diagnose’ patients are excluded. Post hoc and inappropriate exclusions will be labelled 'high risk' of bias.* |
| **Index Test** | | |
| Was ¹⁸F-FDG PET assessment/ interpretation performed without knowledge of clinical dementia diagnosis? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *Terms such as “blinded” or “independently and without knowledge of” are sufficient and full details of the blinding procedure are not required. Interpretation of the results of the index test may be influenced by knowledge of the results of reference standard. If the index test is always interpreted prior to the reference standard then the person interpreting the index test cannot be aware of the results of the reference standard and so this item could be rated as ‘yes’.* |
| Were ¹⁸F-FDG PET thresholds pre-specified? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *For biomarkers there is often a reference point (in units or categories) above which participants are classified as 'test positive'; this may be referred to as threshold; clinical cut-off or dichotomization point. A study is classified at high risk of bias if the authors define the optimal cut-off* ***post hoc*** *based on their own study data, because selecting the threshold to maximize sensitivity and specificity may lead to overoptimistic measures of test performance.* |
| **Reference Standard** | | |
| Is the assessment used for clinical diagnosis of dementia acceptable? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *Commonly-used international criteria to assist with clinical diagnosis of dementia include those detailed in DSM-IV and ICD-10. Criteria specific to dementia subtypes include, but are not limited to, NINCDS-ADRDA criteria for Alzheimer’s dementia; McKeith criteria for Lewy Body dementia; Lund criteria for frontotemporal dementia; and the NINDS-AIREN criteria for vascular dementia. Where the criteria used for assessment are not familiar to the review authors or the Cochrane Dementia and Cognitive Improvement group this item should be classified as 'high risk of bias'; where the criteria are not reported it should be classified as ‘unclear’*. |
| Was clinical assessment for dementia performed without knowledge of the ¹⁸F-FDG PET? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *Terms such as “blinded” or “independently and without knowledge of” are sufficient and full details of the blinding procedure are not required. Interpretation of the results of the reference standard may be influenced by knowledge of the results of index test.* |
| **Participant flow** |  |  |
| Was there an appropriate interval between ¹⁸F-FDG PET and clinical dementia assessment? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *As we test the accuracy of the* ¹⁸F-FDG PET biomarker *for MCI conversion to dementia, there will always be a delay between the index test and the reference standard assessments. The time between reference standard and index test will influence the accuracy (*Geslani 2005*;* Okello 2009*;* Visser 2006*), and therefore we will note time as a separate variable (both within and between studies) and will test its influence on the diagnostic accuracy.* *In this review one year or more is considered to be an appropriate interval period of follow-up for diagnostic verification.* |
| Did all participants get the same assessment for dementia regardless of ¹⁸F-FDG PET? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *There may be scenarios where participants who score “test positive” on index test have a more detailed assessment. Where dementia assessment differs between participants this should be classified as high risk of bias.* |
| Were all participants who received ¹⁸F-FDG PET assessment included in the final analysis? | No=high risk of bias  Yes=low risk of bias  Unclear=unclear risk of bias | *If the number of participants enrolled differs from the number of participants included in the* 2 x 2 *table then there is the potential for bias. If participants lost to drop-out differ systematically from those who remain, then estimates of test performance may differ.* *If there are drop-outs they should be accounted for.* |
| **Anchoring statements to assist with assessment for applicability** | | |
| **Patient Selection** | | |
| Were included participants representative of the general population of interest? | High applicability concerns  Low applicability concerns  Unclear applicability concerns | *The included participants should match the intended population as described in the review question. The review authors should consider population in terms of symptoms; pre-testing; potential disease prevalence; setting. If there is a clear ground for suspecting an unrepresentative spectrum the item should be rated poor applicability.* |
| **Index Test** | | |
| Were sufficient data on ¹⁸F-FDG PET application given for the test to be repeated in an independent study? | High applicability concerns  Low applicability concerns  Unclear applicability concerns | *Variation in technology, test execution, and test interpretation may affect estimate of accuracy. In addition, the background, and training/expertise of the assessor should be reported and taken into consideration. If* ¹⁸F-FDG PET *was not performed consistently this item should be rated poor applicability.* |
| **Reference Standard** | | |
| Was clinical diagnosis of dementia made in a manner similar to current clinical practice? | High applicability concerns  Low applicability concerns  Unclear applicability concerns | *Where the criteria used for assessment are not familiar to the review authors or the Cochrane Dementia and Cognitive Improvement group this item should be classified as poor applicability.* |

**QUADAS 2 assessment and judgment about risk of bias/ Characteristics of studies included in the exploratory analysis**

|  |  |  |
| --- | --- | --- |
| **Anchisi 2005** | | |
| ***Patient selection*** | | |
| **A. Risk of bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort  Consecutive sample of 67 right-handed participants with mild cognitive impairment (Dr Perani’s email on 22/10/2013) and 41 healthy controls. We only included data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable.  Exclusion criteria: depression and behavioral disorders. No further information. |
| Was a consecutive or random sample of patients involved? Yes  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 67 participants with MCI, diagnosed with the Mayo Clinical criteria (Petersen 2001) at baseline, were recruited from 4 centers enrolled in the Network for Efficiency and Standardization of Dementia Diagnosis Fifth European Framework Research Project.  48 participants were assessed at follow-up  Gender: total sample 34 men and 33 women; MCI-non-converters: 20 M, 14 F; MCI-converters: 5 M; 9 F  Age (y): total sample mean 67.7± 8.3; MCI-non-converters: 65.0± 9.0; MCI-converters: 71.1 ± 3.9  APOE ɛ4: not reported  MMSE: mean: total sample 27.7 ±1.7; MCI-non-converters: 28.4 ± 1.1; MCI-converters: 26.6 ± 1.7;  Education (y): total sample mean 11.0 ± 4.7; MCI-non-converters: 11.2 ± 4.5; MCI-converters: 9.1 ± 5.0  Sources of referral: primary care physicians  Sources of recruitment: outpatients from 4 University Departments (Milan, Brescia, Cologne and Dresden) |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| ¹⁸F-FDG PET scan *Acquisition*  Studies were performed according to previously described methods (Herholz 2002). The software packages **SPM**99 (Welcome Department of Cognitive Neurology, University College, London, England) and MATLAB 6.1 (MathWorks Inc, Sherborn, Mass) were used for image pre-processing. Images were spatially normalized to a reference stereotactic template (Montreal Neurological Institute, McGill University, Montreal, Quebec) by a 12-parameter transformation and smoothed by a Gaussian kernel of 12x12x12-mm voxels full width at half maximum. *Data analysis*  The hypometabolic regions in participants with MCI who developed Alzheimer's disease compared with controls, obtained by SPM99 analysis, were used to define volume of interest (VOI). Using only clusters > 700 voxels, 3 VOIs in the temporo-parietal regions and posterior cingulate cortex were selected. The regional sensorimotor ¹⁸F-FDG uptake ratio was used as the index test. Sensitivity and specificity data were reported for a threshold of 1.138, which was derived from ROC analysis.  Threshold: rCGM (regional cerebral glucose metabolism) = 1.138; not pre-specified  At baseline 67 MCI. A number of ‘positive’ and ‘negative‘ tests reported only for 48 MCI participants who had follow-up data: 19 with positive ¹⁸F-FDG test (≤ 1.138); 29 with negative ¹⁸F-FDG test (> 1.138)  Not reported whether the index test results were interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: NINCDS-ADRDA criteria  Unclear whether clinicians conducting follow-up were aware of the ¹⁸F-FDG PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: median follow-up 12 months; range: 12 - 27 months  At baseline 67 MCI.  At follow-up: 48 participants: 14 MCI-ADD; 34 MCI-MCI (p 1730)  *Conversion to AD*  Sensitivity: 93%; Specificity: 82%; NPV: 96%; PPV: 68% (at the threshold of rCGM-r = 1.138; p1731)  Number included in analysis: 48  TP = 13; FP = 6; FN = 1; TN = 28 (calculated in RevMan5)  Loss to follow-up: 19; no further information. |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  We contacted the trial investigators who provided some additional data for the 'Patient selection' and 'Patient characteristics and setting' items (email on 22/10/2013). | | |
|  | | |
| **Arnaiz 2001** | | |
| ***Patient selection*** | | |
| **A. Risk of bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort  Consecutive sample of 20 participants with MCI was recruited from the Geriatric Clinic, Huddinge University Hospital, Sweden  Exclusion criteria: not reported |
| Was a consecutive or random sample of patients involved? Yes  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 20 participants with MCI, diagnosed with the Global Deterioration Scale (Reisberg 1982) at baseline. These criteria were not pre-specified in the protocol.  Gender: Total: 8 women, 11 men; converters: 5F, 6M; non-converters: 3F, 6M  Age (y): converters 64.9 ± 8.3 years; non-converters: 60.1 ± 8.4  APOE ɛ4: not reported  MMSE: converters 26.7 ± 1.8; non-converters: 27.2 ± 2.9  Education (y): 11.9 ± 2.2 years; non-converters: 11.3 ± 2.0  Sources of referral: not reported  Sources of recruitment: Geriatric University Hospital Clinic |
| **Are there concerns that the included patients and setting do not much the review question? Unclear concerns** | | |
| ***Index test*** | | |
| ¹⁸F-FDG PET scan  *Acquisition*  The PET investigations were performed at the Uppsala University PET Center, using either of 2 scanners (GEMS 2048-15B or GEMS 4096-15WB, General Electric Medical Systems, Milwaukee, WI). The accumulation of 2-[18F]-fuoro-deoxyglucose (¹⁸F-DFG) in the brain was followed for 60 minutes.  *Data analysis*  Regions of interest (ROIs) were defined on transaxial slices in relation to the slice where the basal ganglia (BG) structures were best visible. Based on Herholz 1999 and Jelic 1999, rCGM were obtained for 3 ROI: the temporo-parietal regions 13 mm above the level of the basal ganglia (**TPabove**), 13 mm below (**TPbelow),** and at the level of the basal ganglia (**TP BG)** in the left and the right hemispheres. Estimates of the rCMGlu were standardized to the sensorimotor area of the cortex 26 mm above the level of the basal ganglia. This region is thought to be relatively unchanged in people with AD (Duara 1986). The rate of glucose consumption in the brain was expressed in mol/min 3100 cm3 and calculated by a graphical method which used the lumped constant equal to 0.418 for correction of differences in utilization between ¹⁸F-FDG and glucose.  Threshold: visual inspection: **baseline** rCGM of left temporo-parietal region above the basal ganglia (Model I); not pre-specified  Not reported whether the index test results were interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: NINCDS-ADRDA and DSM-IV criteria  Unclear whether clinicians conducting follow-up were aware of the ¹⁸F-FDG PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: average interval 36.5 months  Information from the paper  20 MCI: 9 MCI-ADD; 11 non-converters; baseline rCGM of left TPabove (isolated)  When model I was used (left TPabove measure isolated), the model reached a 75% classification accuracy (P = 0.05). Three participants with pMCI were classified as sMCI, and two sMCI were classified as pMCI (p 853); therefore there were FN = 3; FP = 2  Calculated in Review Manager 5: TP = 6; TN = 9; sensitivity = 67%; specificity = 82%  Number included in analysis: 20  TP = 6; FP = 2; FN = 3; TN = 9  Loss to follow-up: none |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risks** | | |
| **Notes** | | |
|  | | |
| **Berent 1999** | | |
| ***Patient selection*** | | |
| **A. Risk of bias** | | |
| **Patient sampling** | | Study design: nested case-control study; retrospective analysis of the longitudinal data  45 participants were recruited: 18 with AD, 20 with isolated memory impairment (IMI) and 15 healthy volunteers.  Although a sampling procedure was not described, it looks as if a consecutive sample of participants was not recruited; at baseline there were selected 10 IMI participants with positive and 10 IMI participants with negative ¹⁸F-FDG test.  We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable.  Exclusion criteria: no participants or control subjects were taking any centrally-acting medications at the time of study. No further information |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 20 participants with IMI. Participants were screened by staff of the Michigan Alzheimer’s Disease Research Center (MADRC) and classified using the clinical and psychometric IMI criteria: objective and quantitative evidence of learning inefficiency, with no evidence of impairments in general cognitive status or activities of daily living or behavior due to change in cognition. This classification is based largely on previously published **AAMI** criteria (Crook 1986), although the IMI criteria do not require a formal memory complaint, and there is a liberal age restriction.  Gender: 7 women; 13 men  Age (y): mean 70.2 ± 5.5  APOE ɛ4: not reported  MMSE: 26.0 ± 1.9  Education (y): total sample average: 15  Sources of referral: not reported  Sources of recruitment: Cognitive Disorders Clinic, Department of Neurology at the University of Michigan |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| ¹⁸F-FDG PET scan  *Acquisition*  ¹⁸F-FDG PET image sets were acquired following intravenous administration of 10 mCi (370 MBq).  *Data analysis*  Image sets were analyzed in quantitative and non-quantitative (normalization) fashions described elsewhere (Minoshima 1995). Regional glucose metabolism in frontal, temporal, parietal and occipital regions normalized to the thalamus were determined for IMI participants. **3D-SSP** metric used.  Threshold: a diagnostic index based on **Z-scores** of the parietal cortex was used to categorize people with IMI into normal and abnormal rCGM (Minoshima 1995); not pre-specified  Not reported whether the index test results were interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: NINCDS-ADRDA; ICD-10. All participants received both reference standards.  Unclear whether clinicians conducting follow-up were aware of the ¹⁸F-FDG PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 3 years  At baseline: 20 IMI; 10 IMI with positive ¹⁸F-FDG test; 10 IMI with negative ¹⁸F-FDG test.  At follow-up: 10 IMI with positive ¹⁸F-FDG test: 7 IMI-ADD; 3 IMI-IMI; 10 IMI with negative ¹⁸F-FDG test: 3 IMI-ADD; 7 IMI-IMI  Number included in analysis: 20  TP = 7; FP = 3; FN = 3; TN = 7; sensitivity=70%; specificity=70%  Loss to follow-up: none |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes** | | |
|  | | |
| **Bruck 2013** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: nested case-control study; prospective longitudinal data analyzed  Sample and inclusion criteria described in details in Koivunen 2011: 29 MCI participants were initially consecutively recruited; 22/29 MCI participants had undergone baseline 18F-FDG PET scans. Sample procedure for that sample was not reported.  Twenty-nine consecutive MCI patients seen at the memory clinic and who volunteered for PET scanning and thirteen healthy controls were included in the study. In this review we only included data on performance of the index test to discriminate between 22 MCI participants who convert to dementia and those who remained stable, and who had baseline 18F-FDG PET scans.  Exclusion criteria: not reported |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 29 MCI participants diagnosed by Petersen 2004 criteria: all patients had subjective memory impairment that was confirmed in neuropsychological testing, and some patients had mild decline in other cognitive domains. Clinical Dementia Rating (CDR) was 0.5, global cognition was normal, activities of daily living were not impaired, and no subject had dementia at baseline  Gender: 11 women, 18 men  Age (y): total sample 71.7±7.5 years; MCI converters 71.6±7.2 years; non-MCI converters 70.8±4.9  APOE ε4 carrier: not reported  MMSE: total sample 26.9±1.6; MCI converters 26.2±1.5; non-MCI converters 27.9±1.3  Education (y): not reported  Sources of referral: not reported  Setting: secondary care – outpatient memory clinic |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  248 ±47.5 MBq [18F]FDG was injected and a 55-min dynamic scan was performed.  *Data analysis*  Automated region of interest (ROI) analysis was performed; parametric images were calculated. ROIs were defined using Imadeus (version 1.50; Forima, Turku, Finland) on a spatially normalized MRI template image representing brain anatomy. Retention detecting ROI included: lateral frontal cortex (LFC), medial frontal cortex, parietal cortex, lateral temporal cortex (LTC), medial temporal cortex, occipital cortex, anterior cingulate, posterior cingulate/precuneus, caudate and putamen. The reference ROIs were drawn on the pons. The region-to-pons ratio of the FDG concentration over 30–55 min was calculated.  Threshold: 1.16 for [18F]FDG retention in the lateral frontal cortex; derived from the ROC analysis; not pre-specified  Not reported whether the index test results were interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| Target condition and reference standard(s) | | Target condition: conversion from MCI to Alzheimer’s disease dementia  Reference standard: NINCDS-ADRDA; DSM-IV criteria  Note:All participants received the both reference standards |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 2 years  All 22 MCI participants who had 18F-FDG PET scan at baseline were included in the analysis. It was reported that “the ROC analysis showed a cut point of 1.16 for 18F-FDG PET, which resulted in 78% specificity and **87%** sensitivity (p1570).  Number included in analysis (N=22)  13 MCI-AD (disease positive); 9 MCI-MCI (disease negative) with [18F]FDG PET scan at follow-up, based on the information on p1570 (above)  Conversion from MCI to ADD  TP=11; FP=2; FN=2; TN=7 (Calculated in Revman5)  Sensitivity=**85%;** specificity=78% (Calculated in Revman5)  All participants who had undergone baseline 18F-FDG PET scan were included in the analysis. |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes**  There is discrepancy between the sensitivity values on p1570 in the study and our calculation in RevMan5. Author contacted and confirmed (Dr Bruck, email on that 25/3/16) that 22 MCI participants were included in the analysis: 13 MCI-AD and 9 MCI-MCI; included in the analysis; sensitivity=85% (not 87% as reported on p1570). | | |
|  | | |
| **Cerami 2015** | | |
| ***Patient selection*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort  MCI participants were recruited from the San Raffaele Scientific Institute. Sampling procedure not described.  Exclusion criteria: significant positive (e.g., aggressiveness, disinhibition or psychotic disorder) or negative (e.g., loss of empathy or sympathy) behavioral changes. |
| Was a consecutive or random sample of patients involved? Unclear  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Unclear** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 45 participants diagnosed by Petersen criteria (Petersen 2009; Albert 2011).  Gender: total sample 26 women, 19 men  Age (y): total sample 70.5±5.7  APOE ε4 carrier:  MMSE: 26.7±1.9 years  Education (y): mean 1±5.7  Sources of referral: not reported  Setting: The San Raffaele Scientific Institute, Milan, Italy (inpatients) |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  All 18F-FDG PET acquisitions were performed for diagnostic research purposes at the Nuclear Medicine Unit, San Raffaele Scientific Institute, Milan, following standardized procedures (Anchisi 2005) within 6 months from the first baseline clinical visit. Injection of 18F-FDG PET: 185–250Mbq: usually, 5–8 mCi via a venous cannula.  *Data analysis*  Image pre-processing and statistical analysis were performed using **SPM** (http://www.fil.ion.ucl.ac.uk/spm/software/spm5). Single patient 18F-FDG PET scans were normalized according to the procedure implemented in the new standardized SPM5 18F-FDG PET dementia-specific template (Della Rosa et al., 2014) for spatial normalization of 18F-FDG PET scans. This is an optimized method that showed increased reliability and accuracy of estimated metabolic activity patterns compared to the standard [15O]H2O-PET template currently available for SPM normalization procedures. Indeed, it allows to better recognize relevant metabolic changes that may otherwise be obscured by spatial normalization. Each patient scan was tested for relative ‘hypometabolism’ by comparison with a normal reference 18F-FDG PET dataset ad hoc developed and validated by Perani et al., 2014.  Threshold: it was set at p=0.05, FWE-corrected for multiple comparisons at the voxel level. Only clusters containing more than 100 voxels were deemed to be significant; pre-specified  Visual interpretation of 18F-FDG PET scans (**SPM-t maps**) by two expert raters. Each patient scan was labelled as: ‘negative’ or ‘AD-like’ or ‘bvFTD-like’ or ‘sv-PPA-like’ or ‘CBD-like’ or ‘DLB-like’ or ‘mTLD-like’ (p190).  All 18F-FDG PET scans were retrospectively evaluated and independently classified by two expert raters. Both raters were blinded to baseline clinical-neuropsychological data and to the results of reference standards at follow-up. | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| Target condition and reference standard(s) | | Target condition: conversion from MCI to all types of dementia  Reference standards: NINCDS-ADRDA (McKhann 2011) for Alzheimer’s disease dementia; McKeith 2006 for LBD; Rascovsky 2011 for bvFTD  The clinicians were blinded to index test data during all follow-up periods. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 28.5±7.8 months  At baseline: 14 MCI with 18F-FDG PET pattern ‘*negative’*; 31 MCI with 18F-FDG PET pattern *‘positive’* (15 AD-like; 7 FTLD-like; 2 DLB-like; 7 mTLD-like) (Table 2, p189)  Number included in analysis (N=38) (Dr Cerami’s email on 8/8/16)  Conversion from MCI to ‘all dementia’ (‘dementia subtypes’)  TP=19; FP=5; FN=0; TN=14 (Dr Cerami’s email on 8/8/16)  Sensitivity=100% specificity=74% ; PPV=79%; NPV=100% (Calculated in RevMan5)  Number excluded from the analysis: 7 MCI participants with mTLD-like pattern at baseline; at follow-up 2 MCI-AD-converters; 5 MCI-non-converters (Table 2, p189) |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  Dr Cerami’s email on 8/8/16: “We consider MCI patients with mTLD pattern as positive since they showed a pattern of neurodegeneration in the hippocampal structures mirroring the cognitive disorders they presented…we do not know if it is convenient to include MCI cases with FDG-PET evidence of focal medial temporal lobe dysfunction (mTLD, see for example Marra et al., 2012) in your analysis of FDG-PET accuracy in early detecting dementia condition. This is a relatively unknown clinical condition that seems to stay stable for years. The ‘not-conversion’ in such cases may thus not be a false positivity but rather a peculiar feature of this cognitive syndrome. This is the reason why we do not calculated sensitivity and specificity of the whole sample but rather the probability of progression to a "targeted dementia" subtype (AD, FTLD and DLB) in individuals after the evidence of a specific pattern of hypometabolism. Which dementia subtype represents the mTLD pattern we still don't know” | | |
|  | | |
| **Chetelat 2003** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: nested case-control study; prospective longitudinal data analyzed  19 right-handed participants with a memory complaint, but preserved activities of daily living were prospectively recruited. In addition 15 healthy controls were included. Sampling procedure not described. We only included data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable.  Exclusion criteria: neurologic, medical, or psychiatric disorder. |
| Was a consecutive or random sample of patients involved? Unclear  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patients characteristics and settings** | | 19 participants with MCI diagnosed with the Petersen 2001 criteria. Inclusion criteria: no neurologic, medical, or psychiatric disorder; modified Hachinski score ≥ 2; age > 55 years; education > 7 years; episodic memory performance > 1.5 SD below age-matched normal mean in Rey Figure delayed recall or 1 subscore of Grober-Buschke test; Neurological and Communicative Disorders and Stroke, Alzheimer's Disease and related Disorders Association criteria for Alzheimer's disease not met; MMSE ≥ 24 and normal cognitive functions apart from episodic memory, including the Stroop test, visuospatial function, imitation and production of gestures, and language.  Demographic characteristics reported on 17 MCI participants, who had a follow-up.  Gender: 8 men; 9 women. MCI-non-converters: 5M, 5F; MCI-converters: 3M, 4F  Age (y): Total: mean 69.9 ± 6.7; MCI-non-converters: mean 67.8 ± 7; MCI-converters: mean 73 ± 5.1  APOE ε4: not reported  MMSE: ≥ 24 (no further details)  Education (y): not reported  Sources of referral: not reported  Sources of recruitment: not reported |
| **Are there concerns that the included patients and setting do not much the review question? Unclear concerns** | | |
| ***Index test*** | | |
| ¹⁸F-FDG PET scan  *Acquisition* At entry each participant underwent an ¹⁸F-FDG PET scan using the ECAT HR+ device (CTI, Knoxville, TN).  Data analysis  The ¹⁸F-FDG uptake datasets were handled with **SPM**99. SPM maps were threshold at Z > 3.09; only decreases were assessed. Two cerebral regions were mainly evaluated: the right temporo-parietal and posterior cingulate. The participants were classified according to the adjusted regional activity values in the referred areas.  Threshold: not pre-specified: thresholding was set at 80% of whole brain mean of control participants. Not reported the index test results were interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: NINCDS-ADRDA criteria  Clinicians conducting follow-up were blinded to the ¹⁸F-FDG PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 18 months. Participants were evaluated every 6 months for an 18-month period  At baseline 19 MCI.  At follow-up: 17 participants: 7 rapid converters (MCI-ADD); 10 non-converters (MCI-MCI) (p 1377)  Number included in analysis: 17  TP = 7; FP = 0; FN = 0; TN = 10 (right temporo-parietal region) (Figure, p 1376); sensitivity=100%; specificity=100%  TP = 7; FP = 1; FN = 0; TN = 9 (posterior cingulate region) (Figure, p 1376); sensitivity=100%; specificity=90%  Loss to follow-up: 2 participants were excluded post hoc: 1 refused repetitive cognitive testing, and another turned out to have depression (did not meet inclusion criteria) |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? Unclear risk** | | |
| **Notes**  We included data for the temporo-parietal region in the exploratory analysis, since it represents a typical and 'wider' brain area that is potentially involved in conversion to AD dementia. | | |
|  | | |
| **Choo 2013** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort  83 MCI patients were recruited from the Department of Geriatric Medicine of memory disorders. Sampling procedure: consecutive recruitment (Dr Nordberg’s email on 21/3/2016)  Exclusion criteria: not reported |
| Was a consecutive or random sample of patients involved? Yes  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 83 MCI participants diagnosed by Petersen 1999 criteria were included in the study. Demographic characteristics reported on 77 MCI participants who were included in the analysis.  Gender: 18 women and 8 men MCI converters; 25 women and 26 men MCI-non-converters  Age (y): MCI converters 63.5±8.2; non-MCI converters 59±7.5  APOE ε4 carrier: 19/26 (72%) MCI converters; 23/51 (56%) MCI-non-converters  MMSE: MCI converters 26.6±2.6; non-MCI converters 28.5±1.6  Education (y): not reported  Sources of referral: the primary care centers in the community for investigation of suspected dementia development.  Setting: Department of Geriatric Medicine, Karolinska University Hospital Huddinge, Stockholm, Sweden |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  All subjects underwent PET examinations with 18F-FDG at Uppsala PET Centre/Imanet. PET examinations were performed under resting conditions, in dimmed light room, and with the patients’ head suitably fixed to prevent movement and fasting before for 4 h. Accumulation of FDG in the brain was monitored for 60 min, during which 13 samples were taken to examine variations in FDG concentration. The cerebral metabolic rate of glucose (CMRglc), expressed in mol/min/100 g, was calculated by a graphical procedure [17], including a lumped constant of 0.418 to correct for differences in utilization between FDG and glucose.  *Data analysis*  Five ROIs (frontal, temporal, parietal, posterior cingulate cortex (PCC), caudate) were used for regional glucose metabolism analysis, which were mean of each compatible right and left, 10 ROIs selected from each manually drawn 46 ROIs of GE scanner for 76 MCI subjects and 24 ROIs of ECAT EXACT HR+ scanner analyzed by Hammers atlas from additional 8 MCI cases. Composite ROI was also calculated by mean of aforementioned 5 ROIs. **SPM** used for image analysis.  Threshold: 1.505; parietal glucose metabolic rate by pons normalization (Dr Nordberg’s email on 21/3/2016); not pre-specified  Unclear the index test results interpreted without knowledge of the results of the reference standard. | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: Conversion from MCI to Alzheimer’s disease.  Reference standard: NINCDS-ADRDA and DSM IV criteria for Alzheimer’s disease dementia. Reference standards for non-AD dementias not reported.  Researchers were blinded to PET results. |
| Is the reference standard likely to correctly classified the target condition? Unclear  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: mean 44±35.4 months (range, 1.6–161.7 months)  At follow-up: 32 MCI-converters: **26 MCI-AD dementia** and 6 MCI-non-AD dementia (1 MCI-FTLD; 4 MCI-vascular dementia; 1 MCI-Parkinson-dementia); **51 MCI-non-converters** (MCI-MCI); 4MCI-normal cognition  Number included in analysis (N=77)  Conversion from MCI to ADD  Sensitivity=77%; specificity=80% (Dr Nordberg’s email 18/3/16)  TP=20; FP=10; FN=6; TN=41; PPV=66.7%; NPV=87% ( (calculated in RevMan5)  Note: six MCI participants who were converted to non-AD dementia were excluded from the analyses, because progression to AD was the main focus of this study. |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Unclear  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  Dr Nordberg’s email on 21/3/2016: consecutive sample, collected from clinical patient investigation of memory disorders; Threshold: 1.505 (Parietal glucose metabolic rate by pons normalization).  The authors explored the best combination model with baseline demography, neuropsychology, 18F-fluorodeoxyglucose positron emission tomography (FDG-PET), cerebrospinal fluid (CSF) biomarkers, and apolipoprotein E (APOE) genotype evaluation to predict progression to AD in mild cognitive impairment (MCI) patients.  Conclusion: Findings highlight that the combination of regional glucose metabolic assessment by PET and CSF biomarkers evaluation can significantly improve AD predictive diagnostic accuracy of each respective methods | | |
|  | | |
| **Clerici 2009** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: nested case-control study; retrospective analysis of longitudinal data  Sampling procedure not described.  Information from the author: 16 aMCI came from the Del Sole 2008 study; 14 snaMCI were added to the current study.  Exclusion criteria: i) presence of a DSM-IV psychiatric disorder, including dementia or of organic brain pathology or of organic illness affecting the brain; ii) significant history of head injury; iii) major systematic illness; iv) history of drug and alcohol dependence; v) history of stroke. |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 30 MCI (16 aMCI and 14 snaMCI) were included. The diagnostic criteria for MCI were: 1) subjective and objective anamnestic evidence of progressive cognitive impairment for more than 6 months; 2) normal activities of daily living; 3) MMSE score of 24 or greater; 4) a CDR score of 0.5; and 5) a score > 1.5 SD below the mean on at least 1 cognitive dimension, as evaluated by neuropsychological assessment.  Gender: aMCI: 10 women (62.5%) and 6 men (37.5%); snaMCI: 10 women (71.4%) and 4 men (28.6%)  Age (y): aMCI: 74.92 ± 7.6 years; snaMCI: 73.62 ± 6.3  APOE ɛ4: not reported  MMSE: aMCI: 25.82 ± 1.5; snaMCI: 26.72 ± 1.9  Education (y): aMCI: 9.1 ± 4.5 years; snaMCI: 8.7 ± 4.0 years  Sources of referral: GP surgeries or specialists (85%) or self-referral (15%)  Sources of recruitment: Center for Research and Treatment of Cognitive Dysfunctions of the Department of Neurology, University of Milan, Italy |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| ¹⁸F-FDG PET scan  *Acquisition*  An activity of 185-370 MBq of ¹⁸F-FGD, depending on person’s weight, was injected intravenously in resting condition with eyes closed and ears unplugged; the participants were asked to rest quietly for the next 45 minutes. The studies were performed using an ECAT ACCELL scanner (Siemens Medical Systems, Erlangen, Germany).  *Data analysis*  PET data of MCI participants were compared to a control group of 7 cognitively normal elderly participants described in previous study of the group (Del Sole 2008), and then to each other on a voxel-by-voxel basis using a 2-sample t test.  Each PET study was analyzed separately (according to the method described in the Del Sole 2008 study) to assess regional cerebral metabolic abnormalities in individual participants. Briefly, the **SPM(t) maps** of each person were converted to binary masks, where single pixels of the images were either a 0 in areas of normal ¹⁸F-FDG uptake or 1 in areas of decreased uptake. The mask images were summed together to generate a map of overlapping regions of metabolic impairment.  Threshold: Each scan was considered positive when a cluster of at least 100 consecutive voxel (size 2 x 2 x 2 mm³) had a metabolism lower that the control group (with P set at < 0.01 level); pre-specified (Dr. Clerici email on 23rd August 2013). | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer's disease dementia and other forms of dementia  Reference standard: NINCDS-ADRDA and DSM-IV for AD dementia; McKeith criteria for LBD; Lund and Manchester criteria for FTD  Not clear whether clinicians conducting follow-up were aware of the ¹⁸F-FDG PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: aMCI group: last follow-up 18 months; snaMCI group: follow-up at 12, 24, and 37 months  Information from the author:  At baseline: 26 ¹⁸F-FDG+ tests; 4 ¹⁸F-FDG- tests at baseline  At follow-up (37 months):  12 aMCI with ¹⁸F-FDG+: 11 aMCI converters (10 aMCI-ADD; 1aMCI-LBD), 1 lost to follow-up  4 aMCI with ¹⁸F-FDG-: 1aMCI-ADD; 2aMCI-MCI; 1 lost to follow-up  14 snaMCI with ¹⁸F-FDG+: 7 converters (2 snaMCI-ADD; 2 snaMCI–FTD; 3 snaMCI-LBD) and 5 non-converters (5 snaMCI- snaMCI) and 2 lost to follow-up  Number included in analysis: 26  TP = 12; FP = 11; FN = 1; TN = 2 for Alzheimer's disease dementia  Sensitivity=92%; specificity=15%  TP = 18; FP = 5; FN = 1; TN = 2 for all forms of dementia  Sensitivity=95%; specificity=29%  Loss to follow-up: In total 4 MCI participants: 2 aMCI and 2 snaMCI. No further details. |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Unclear  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  We contacted the trial investigators who provided relevant data tor the 2 x 2 table to be completed (email on 23rd August 2013). | | |
|  | | |
| **Drzezga 2005** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort  Sampling: participants were consecutively recruited  Exclusion criteria: people who met the diagnostic criteria for dementia or any other functional psychiatric disorder, including major depression; symptoms of diseases or abnormalities sufficient to cause memory impairment (e.g., Parkinson's disease, normal pressure hydrocephalus); major structural abnormalities on MRI (e.g., infarction, intra-cerebral aneurysm, arteriovenous malformation); extra-cerebral causes which could influence neuropsychological function (e.g., use of neuroleptics, substance abuse).  The study excluded people with depression, but specified major depression sufficient to cause memory impairment. |
| Was a consecutive or random sample of patients involved? Yes  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 30 MCI participants, diagnosed with the Petersen 1999 and CDR=5 criteria, were recruited from a research unit.  Baseline evaluation included medical, psychiatric and neurological examinations performed by an experienced psychiatrist. Participants had to meet the established diagnostic criteria for mild cognitive impairment: subjective complaint; performance of 1.5 SD below the age norm on the Consortium to establish a registry for Alzheimer's Disease (CERAD) delayed verbal recall test; CDR score of 0.5; preserved basic activities of daily living.  Gender(y): 14 men; 16 women; MCI-non-converters: 8M, 10F; MCI-converters: 6M, 6F  Age: mean: total sample 70 ± 8 years; MCI-non-converters: 67.6 ± 8.2 years; MCI-converters: 74.7 ± 4.7 years  APOE ɛ4: MCI-non-converters: 8/18; MCI-converters: 9/12  MMSE: MCI-non-converters: 27.6 ± 1.5; MCI-converters: 25.9 ± 2.1  Duration of symptoms (y): mean 2.6 ± 2.0  Education (y): mean 11.6 ± 3.4 years  Sources of referral: GP surgeries or neurologists or psychiatrists or other institutions  Sources of recruitment: Research Unit for Cognitive Disorders, Technical University, Munich, Germany. |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| ¹⁸F-FDG PET scan  *Acquisition*  The index test was performed at the time of initial clinical evaluation. All participants received 370 MBq ¹⁸F-FDG at rest with the eyes closed; 30 minutes after injection, PET was performed under standard resting condition (eyes closed in dimmed ambient light) using a Siemens 951 R/31 PET scanner (CTI). A sequence of 3 frames of 10 min was started and later combined into a single frame. Image data were acquired in 2-dimensional mode with a total axial field of view of 10.5 cm and no interplane gap space. Attenuation correction was performed by a standard ellipse-fitting method.  *Data analysis*  A well-established observer-independent programme (**NEUROSTAT**; University of Michigan) was used to minimize observer bias. This method has been evaluated for clinical and scientific use in people with dementia and other cerebral disorders (Bartenstein 1997; Drzezga 1999; Ishii 2001; Minoshima 1995). The ROIs were defined to reflect functional divisions of the cerebral lobes, and each hemisphere was divided into the following regions: orbitofrontal, prefrontal, premotor, central, parietal superior and inferior, occipital, temporal anterior, temporal posterior and posterior cingulate. The results from the ROI analysis were not averaged together; each ROI was assessed individually.  The detection of significant hypometabolism (as compared with a control population) in surface ROIs covering the posterior cingulate cortex accompanied by cortical hypometabolism in at least unilateral temporo-parietal areas was determined as suggestive of early AD, based on findings of earlier studies (Drzezga 2003). According to this strategy, PET **baseline** results were classified as suggestive or not suggestive for AD. **3D-SSP** of the z-scores was generated to allow visualization of abnormality.  Threshold: A ***z*-score** threshold of > 1.64 (1-tail) corresponding to a P value of 0.05 (1-tail) was applied for demarcation of significant abnormalities. This statistical threshold previously proved to be suitable for the diagnosis of DAT using the applied statistical tool (Bartenstein 1997; Minoshima 1995); pre-specified.  PET baseline results were blinded for the later outcome of the patients and blinded for other clinical baseline information. (p1627) | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: NINCDS-ADRDA criteria  ¹⁸F-FDG PET results were blinded for the later outcome of the participants, and blinded for other clinical baseline information. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 15 months (expanded to a mean 16 ± 2 months)  At baseline: 30 participants: 13 with ¹⁸F-FDG positive; 17 with ¹⁸F-FDG negative (Abstract)  At follow-up: 12 MCI-ADD; 18 MCI-MCI (p 1628); sensitivity: 92%; specificity: 89% (Table 2, p 1629)  Number included in analysis: 30  TP = 11; FP = 2; FN = 1; TN = 16; (Calculated in Review Manager 5)  Loss to follow-up: none |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes** | | |
|  | | |
| **Fellgiebel 2007** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort  Participants with aMCI, presenting at a memory clinic for diagnostic evaluation, were recruited.  Sampling procedure not described.  Exclusion criteria: people with metabolic disease that could affect cognitive function; people with other brain diseases; people with a diagnosis of depression according to DSM-IV criteria |
| Was a consecutive or random sample of patients involved? Unclear  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patients characteristics and setting** | | 16 participants, diagnosed with the Petersen 1999 criteria at baseline. 1 person in the initial study group refused further participation and has been replaced by a consecutively-recruited comparable patient from the memory clinic to preserve the statistical power for prospectively planned follow-up analyses.  Gender: 9 men; 7 women.  Age (y): total sample: mean age 68.6 ± 7.9; MCI-MCI: 68.8 ± 10.0; MCI-progressive: 68.5 ± 5.9 (4/8 MCI-ADD: 69.5 ± 7.9)  APOE ɛ4: not reported  MMSE: mean 25.7 ± 2.7; MCI-MCI: 27.3 ± 1.8; MCI-progressive: 25.0 ± 2.1 (4/8 MCI-ADD: 24.3 ± 1.5)  Education (y): not reported  Sources of referral: not reported  Sources of recruitment: University Memory Clinic, Germany |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  Method of the index test administration described previously (Fellgiebel 2004): *“Acquisitions were in the three-dimensional (3D) mode. Thirty minutes after injection of 180 MBq ¹⁸F FDG, a sequence of three 5-minute frames was started and later combined to a single frame. Thereafter, the images were corrected for attenuation, scatter, and dead time”.*  *Data analysis*  Standardized 3-D stereotactic surface projections (3D-SSP) for each participant, compared with a normal database to provide Z scores, were performed.  Threshold(s): AD-typical findings were defined as significant decrease (**Z-score** > 2 in more than 50 adjacent pixels) of cerebral glucose metabolism in at least 1 of the brain regions that have been shown to be typically involved in early AD (parietal mesial or posterior cingulate and temporal regions); pre-specified.  Unclear whether the index test results were interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: Progression to Alzheimer's disease dementia was assumed if CDR reached 1.  Follow-up evaluation at variable time points (not specified), comprising neurological and psychiatric examination, CDR and MMSE.  Progressive cognitive decline was defined as MMSE score reduction ≥ 2 and a clinical judgement of cognitive deterioration.  Clinicians conducting follow-up were blinded to the¹⁸F-FDG PET results. |
| Is the reference standard likely to correctly classified the target condition? Unclear  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Unclear concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 19 months on average; mean 19.6±9.0  At baseline: 16 MCI: 7 with ¹⁸F-FDG positive; 9 with ¹⁸F-FDG negative  At follow-up: 16 MCI: 7 FDG positive: 4 MCI-ADD, 1 MCI-MCI, 2 MCI-progressive (non-converters); 9 FDG-: 7 MCI-MCI ; 2 MCI-progressive (non-converters) (p 170).  Number included in analysis: 16  TP = 4; FP = 3; FN = 0; TN = 9  Sensitivity: 100%; Specificity: 75%; PPV: 57%; NPV: 100% (calculated in Review Manager 5).  Loss to follow-up: 1/16; however, that participant was replaced by an additional, consecutively-recruited patient from the memory clinic. |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes** | | |
|  | | |
| **Galluzzi 2010** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: retrospective analysis from longitudinal cohort data  Initially 108 consecutive participants with MCI, referred to an outpatient memory clinic over 24 months, were selected. 90 participants were included. Of these, only 38 underwent ¹⁸F-FDG PET scan. Sampling procedure not described.  Exclusion criteria: not specified. |
| Was a consecutive or random sample of patients involved? No  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability: Low concerns** | | |
| **Patient characteristics and setting** | | 38 MCI participants with ¹⁸F-FDG scan. Diagnostic criteria for MCI were not directly specified. However, it can be inferred that the authors use the Petersen 1999 criteria. MCI is defined as the presence of objective impairment in memory or other cognitive domains (performance lower than the 5th percentile on neuropsychological tests applied in the study) in the absence of functional impairment. Demographic data reported on all 90 participants included in the study.  Gender: 53 women, 37 men  Age (y): MCI-NC: 70.9 ± 7.1 years; MCI-ADD: 72.2 ± 7.1 years; MCI-non-ADD; 73.0 ± 7.1  APOE ɛ4: MCI-NC: 19 (41%); MCI-ADD: 14 (58%); MCI-nADD: 2 (15%)  MMSE: MCI-NC: 26.3 ± 1.9; MCI-ADD: 26.4 ± 1 .6; MCI-non-ADD: 25.5 ± 1.9  Education (y): MCI-NC: 7.7 ± 3.6; MCI-ADD: 8.8 ± 4.6; MCI-non-ADD: 7.3 ± 4.0  Sources of referral: not reported  Setting: Translational Outpatient Memory Clinic (TOMC), at the National Institute for the Research and Care of Alzheimer’s Disease (IRCCS Centro San Giovanni di Dio Fatebenefratelli), Brescia, Italy |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  ¹⁸F-FDG PET imaging of the brain was performed at the Nuclear Medicine Service, Spedali Civili of Brescia, by 24-ring 3D PET/CT device (Frisoni 2007)  *Data analysis*  “FDG uptake was assessed with the automated version (PALZ score of PMOD technologies) of the t sum score developed by Herholz and colleagues for the diagnosis of AD, combining the virtues of voxel-based parametric mapping with the diagnostic information on brain regions that are typically affected in AD. Briefly, the ¹⁸F-FDG PET image of an individual patient is compared to a database of normal controls and the voxel-by-voxel sum of t scores in an AD-pattern mask is computed. Abnormal ¹⁸F-FDG PET was defined following the original indications of a t sum higher than 11,090”. (p 2007).  Threshold: **t sum** > 11.090 (¹⁸F-FDG PET positive) (Herholz 2002); pre-specified.  Unclear the index test results were interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standards** | | Target condition: conversion from MCI to Alzheimer's disease dementia and other forms of dementia  Reference standard: NINCDS-ADRDA criteria; DSM-IV; McKeith 1996 criteria; McKhann 2001 criteria  Clinicians conducting follow-up were blinded to the ¹⁸F-FDG PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: MCI-NC: 26.5 ± 16.0 months; MCI-ADD: 21.5 ± 10.2 months; MCI-non-ADD: 19.1 ± 8.9 months  The data refer to all 90 participants with MCI. All 37 MCI participants who had baseline ¹⁸F-FDG scan were included in the analysis.  Information from the author:  At baseline: 28 ¹⁸F-FDG test positive; 10 ¹⁸F-FDG negative  At follow-up: 28 with abnormal ¹⁸F-FDG PET scan: 15 MCI-converters (11 MCI-ADD; 4 MCI non-ADD) and 13 MCI-non-converters (13 MCI-MCI); 10 with normal ¹⁸F-FDG PET scan: 5 MCI-converters (3 MCI-ADD; 2 MCI-non-ADD) and 5 MCI-non-converters (5 MCI-MCI).  Number included in analysis: 38  TP=15; FP=13; FN=5; TN= 5 (conversion to All dementia); sensitivity=75%; specificity=28%  TP=11; FP=17; FN=3; TN=7 (conversion to ADD); sensitivity=79%; specificity=29%  TP=4; FP=24; FN=2; TN=8 (conversion to non-ADD dementia); sensitivity=67%; specificity=25%  Loss to follow-up: none of 38 MCI participants with ¹⁸F-FDG scan  Lost to follow-up for the initial sample: 52 (25 participants refuse the ¹⁸F-FDG PET scan; 7 were not performed because of contraindications and 20 because they had previously undergone 99mTc-ECDSPECT scan).  In addition, 18 participants were excluded from the consecutive sample (N = 108): 16 due to refusal of follow-up; 2 due to logistical problems. |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes**  We contacted the trial investigators who provided relevant data tor the 2 x 2 table to be completed (email on 23/8/2013) | | |
|  | | |
| **Grimmer 2016** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: retrospective analysis from longitudinal cohort data.  Sampling: MCI outpatients who underwent brain MRI, including 3-dimensional dataset, and 18F-FDG and 11C-PIB PET scans, and for whom clinical follow-up data after 24 months were available were selected and included in the analysis.  Exclusion criteria: patients with brain diseases such as normal-pressure hydrocephalus, brain tumors, or Parkinson disease; treatment with psychotropic drugs; addiction to alcohol or legal or illegal drugs; abnormal routine laboratory results; and symptoms indicative of other disorders that might lead to cognitive impairment, for example, sleep apnea, were excluded from the analyses. |
| Was a consecutive or random sample of patients involved? No  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patients characteristics and settings** | | 28 MCI participants, diagnosed with Winbald criteria (revised Petersen criteria): 12 MCI-converters (9 MCI-AD; 2 MCI-FTD; 1 MCI-dementia unknown) and 16 MCI-stable (16MCI-MCI).  Gender: 14 women, 14 men  Age (y): 61.7±7.3 (range 50-78)  APOE ε4 carrier: not reported  MMSE: not reported  Education (y): not reported  Sources of referral: referred by general practitioners, neurologists, psychiatrists, or other institutions, or were self-referred  Setting: Centre for Cognitive Disorders, Department of Psychiatry and Psychotherapy, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| [18F]FDG PET scan  *Acquisition*  18F-FDG PET scans followed standardized protocols. 18F-FDG images were co-registered to high resolution 3-dimensional MRI scans.  *Data analyses:*  Fully automated image analysis: a validated software program, the PMOD Alzheimer Discrimination Tool (**PMOD** Technologies Ltd.), was used. This software classifies 18F-FDG images as either normal or abnormal for AD depending on the extent of voxels showing decreased tracer uptake in AD-typical regions. **‘t-sum’** image analysis used. Cut-off= **11,089 voxels,** p206; pre-specified  Computer aided visual analysis: As in daily clinical routine, the original color-coded axial images, surface projections, and surface projections of difference images compared with a normative dataset using the **Neurostat** routine were provided. Images had to be classified as either typical of AD or not typical of AD. In addition, 18F-FDG images that were rated as not typical of AD were sub-classified in terms of whether they were typical of other neurodegenerative diseases (e.g., frontotemporal lobar degeneration or Lewy body dementia). To be able to compare the results of visual and fully automated analyses, visual analysis findings classified as typical of AD were categorized as positive for AD, whereas all other results were categorized as negative for AD.  The visual analysis was conducted by two independent and board-certified specialists in nuclear medicine who were blinded to the clinical information and the results of the automated analyses. | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standards** | | 1. Target condition: Conversion from MCI to Alzheimer’s dementia and other forms of dementia 2. Reference standard: NINCDS-ADRDA (McKhann 2011); The Lund 1994 for FTD   Clinicians were blinded to baseline MCI subtype and for PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 24 (31.2±7.8) months  Conversion from MCI to Alzheimer’s disease dementia (n=28)  Computer aided visual Analysis  Rater 1: TP=1/28, FP=4/28, FN=8/28, TN=15/28; Sensitivity=11%, Specificity=79%; PPV=20.0%, NPV=65%, Accuracy=54%  Rater 2: TP=5, FP=5, FN=4, TN=14; Sensitivity=56%, Specificity=74%; PPV=50.0%, NPV=78%, Accuracy=68%  Fully automated Analysis (prediction of ADD): TP=7/28, FP=12/28, FN=2/28; TN=7/28, Sensitivity=78%, Specificity=37%; PPV=37%, NPV=78%, Accuracy=50%  Conversion from MCI to All dementias (n=28)  Visual Analysis  Rater 1: TP=10/28, FP=7/28, FN=2/28, TN=9/28; Sensitivity=83%, Specificity=56%; PPV=59%, NPV=82%, Accuracy=68%  Rater 2: TP=9/28, FP=7/28, FN=3/28, TN=9/28; Sensitivity=75%, Specificity=56%; PPV=56%, NPV=75%, Accuracy=64% (FDG positive test=16; FDG negative test=12) |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes**  The author contacted and missing data for creating 2X2 table were provided (Dr Grimmer email on 12/5/2016).Accuracy rates after visual analysis were higher (54%–68%) than those achieved after fully automated analysis (50%), thus suggesting that for 18F-FDG images, visual analysis by experts should be regarded as the preferred option.  Data for Neurostat/3D-SSP were considered in the exploratory analysis because the majority of the studies included in this exploratory analysis applied ‘computer aided visual read’. Data from the Rater 2 who provided more accurate estimates of the diagnostic accuracy of Alzheimer's disease type imaging were included as it is very likely that those readers are more experienced in interpreting brain PET scans. | | |
|  | | |
| **Hatashita 2013** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort.  Participants with MCI were recruited from memory clinic. Sample procedure not reported.  Exclusion criteria: not reported |
| Was a consecutive or random sample of patients involved? Unclear  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 68 MCI participants diagnosed by the Core Clinical Criteria for MCI (almost identical to the criteria previously described by Petersen) (Albert 2011); those criteria correspond to previously described Petersen criteria.  Gender: not reported  Age (y): total sample: range=50-89  APOE ε4 carrier: MCI converters 47%; non-MCI converters 37%  MMSE: MCI converters 26.5±1.5; non-MCI converters 27.3±1.6  Education (y): not reported  Sources of referral: not reported  Setting: memory clinic at Shonan-Atsugy Hospital, Japan |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition:* participants were injected intravenously with 249.9 +/- 28.8 MBq (n= 68) of [18F]-FDG. Fifteen-minute static FDG-PET scans were acquired after a 45-minute uptake period. A Siemens ECAT ACCEL scanner in 3-dimensional scanning mode, providing 63 contiguous 2.46-mm slices with a 5.6-mm transaxial and a 5.4-mm axial resolution were used. All imaging data were reconstructed into a 128x128 matrix.  *Data analyses:* a standardized uptake value (SUV) of the same region was obtained and subsequently normalized to the cerebellar cortex as reference. Glucose metabolism was referred to as the SUV ratio (SUVr).  Threshold: FDG SUVr ≤0.99; pre-specified  Unclear the index test results interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer’s disease dementia  Reference standard: NINCDS-ADRDA criteria  Unclear whether the reference standard results interpreted without knowledge of the results of the index test |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 19.2±7.1 months  Number included in analysis (N=68)  Conversion from MCI to ADD  At follow-up: 30 MCI-converters (disease positive); =30; 38 MCI-non-converters (disease negative)  Sensitivity=93%; specificity=24%; PPV=49%; NPV=82% (Table 2, pe668777)  TP=28; FP=29; FN=2; TN=9 (Calculated in Revman5)  Loss to follow-up: none |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes**  **Authors’ conclusion:** *“A positive FDG PET alone or combined positive Abeta and FDG biomarkers did not significantly discriminate MCI due to AD in individual MCI subjects. Although a FDG PET biomarker has been associated with the neuropathology of AD, it is not specific for AD. In contrast, when assessed in the precuneus with regional hypometabolism, both Abeta and FDG PET biomarkers had a greater level of certainty for MCI due to AD with the greatest sensitivity of 96.6% and specificity of 73.3%. The combination had the highest positive predictive value of 87.8%. We suggest that most individuals with a diagnosis of MCI due to AD, using Abeta and FDG PET biomarkers in a cortical region of precuneus, progress to AD even within a short period”.* | | |
|  | | |
| **Iaccarino 2015** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: retrospective analysis from longitudinal cohort data.  Patient recruited from memory clinics. Multicenter study including San Raffaele Hospital/Vita-Salute San Raffaele University, Milan, Italy and Karolinska Institutet, Stockholm, Sweden.  Sampling procedure: participants with a clinical diagnosis of MCI (Petersen et al., 2014 JIM), 18F-FDG PET and amyloid-PET available and presence of a clinical follow-up to evaluate conversion to AD dementia.  Exclusion criteria: diagnosis of AD dementia (McKhann et al., 2011) at baseline, PET scans not available, corrupted data, lack of extensive clinical follow-up |
| Was a consecutive or random sample of patients involved? No  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 37 MCI participants: 20 MCI-converters and 17 MCI-stable. The diagnostic criteria for MCI: Petersen et al., 2014 JIM  Gender: 25 F, 12 M  Age (y): total sample mean 64.65±7.49  APOE ε4 carrier: 20/30 positive for at least one ε4 allele (data available for N=30/37 subjects)  MMSE: 27.11±2.04  Education (y): 13.06±3.72  Sources of referral: memory clinics  Setting: multi-center |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition* and *Data analysis* as previously described (Perani et al., 2014 NeuroImage: Clinical). SPM-t maps, obtained through an optimized single-subject **SPM** analysis were evaluated by four neuroimaging experts (Cohen’s k: 0.89).  Threshold: the authors reported that they did not use quantitative thresholds as regards 18F-FDG PET evaluation; instead they used the voxel-wise SPM procedure as described in Perani et al., 2014 NeuroImage: Clinical. Positivity thus means presence of AD-like or other dementia-like FDG patterns; pre-specified. Raters were blinded to the results of baseline 18F-FDG PET images (‘positive’ or ‘negative’) to the diagnostic classification at the follow-up | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standards*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standards** | | Target condition: conversion from MCI to Alzheimer’s disease dementia  Reference standard: NINCDS-ADRDA criteria  Unclear whether the reference standard results were interpreted without knowledge of the results of the index test. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: median duration 24 months, inter-quartile range 29.5 months  Number included in analysis (N=**37**)  At follow-up:20 MCI-AD (disease positive); 17MCI-MCI (disease negative)  Conversion from MCI to AD dementia  TP=17; FP=1; FN=3; TN=16  Sensitivity=85**%;** specificity=94**%;** PPV=94**%;** NPV=84**%** (Dr Perani’s email on 11/5/16) |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes**  The paper is under review in E J Nuc Med. Dr Perani provided the data for completing data extraction table (email on 11/5/16). | | |
|  | | |
| **Ito 2015** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: prospective longitudinal cohort  Participants with amnestic MCI were recruited between January 2006 and March 2007 from nine memory clinics. Sampling procedure not reported.  Exclusion criteria: significant underlying medical, neurological, or psychiatric illnesses; patients with formal education for less than 6 years were excluded; no further details |
| Was a consecutive or random sample of patients involved? Unclear  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 114 patients with aMCI who fulfil the inclusion criteria based on a battery of neuropsychological tests: an MMSE score ≥24, a GDS score ≤10, a WMS-R LM I score ≤13, a LMII part A and part B score (maximum = 50) ≤8, and a CDR memory box score equal to 0.5. All participants were living independently in the community at the time of their baseline evaluation (participants with normal activities of daily living).  Gender: not reported  Age (y): MCI converters 71.2±6.5; non-MCI converters 70.5±6.7  APOE ε4 carrier: not reported  MMSE: MCI converters 25.6±1.7; non-MCI converters 26.9±2.0  Education (y): not reported  Sources of referral: not reported  Setting: memory clinics of 9 centers specializing in AD and other dementias across Japan (Supplementary Table 1) |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  Prior to baseline 18F-FDG PET scanning, all subjects fasted for at least 4 h. Intravenous administration of 18F-FDG (254±107 MBq) was followed by a resting period of 40–60 min in a dimly-lit and quiet room, where participants were instructed to keep their eyes open. A static scan for 10.3±5.5 min was performed in 2D or 3D mode after the resting period.  *Data analysis*  The 18F-FDG PET images were processed with the 3-dimensional stereotactic surface projections (3D-SSP) technique to generate z-score maps, using iSSP software version 3.5 (Nihon Medi-Physics Co. Ltd., Tokyo, Japan).  Threshold  ***Visual interpretation*:** visual interpretation of standard imaging by three experts.  **Neurostat/3DD-SSP:** the normal database used for generating the **z-score map** was constructed based on 50 normal subjects (31 males and 19 females, average age = 57.6 y), with 10 normal subjects each from 5 participating institutions.  Three experts, **blinded to clinical information**, independently assessed the reconstructed PET images, referring to the **3D-SSP z-score map** and correlating with MRI to classify the images into different dementia patterns of P1-P3, P1+, and N1-N3 (Silverman 2001). When evaluations of the three raters did not completely match, the cases were discussed, and a consensus reading was agreed upon.  ***PET score****:* ˃1.0; pre-specified (duration of follow-up 3 years)  It was hypothesized that participants with a PET score **at baseline** above 1.0 had a significantly increased risk for progression.  The **AD t-sum** was calculated by using the procedure implemented as module PALZ in the PMOD software package (version 3.2; PMOD Technologies, Zurich, Switzerland), and converted into the PET score by reference to its upper limit of normal as determined previously (Herholz 2002), and log transformation to approach a normal distribution of values, according to the following equation (Herholz 2011) PET score = log2 *{*(AD t-sum/11,089)+1*}*.  **PET score=1.03**, obtained using Youden index (duration of follow-up 2 years) | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Low concerns** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standards*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standards** | | Target condition: conversion from MCI to Alzheimer’s disease dementia  Reference standard: both NINCDS-ADRDA criteria and CDR≥1.0  Researchers were blinded to the PET results. |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: 36 months  114 aMCI recruited to the study.  Number included in analysis (N=88)  Conversion from MCI to ADD: at follow-up 88 aMCI: **41** aMCI-AD converters (disease positive) and **47** aMCI non-converters (disease negative)  Duration of follow-up (N=88): 1 year  *Visual interpretation of standard imaging*: sensitivity=100%; specificity=22% (Table 2, p548)  TP=41; FP=37; FN=0; TN=10 (Calculated in Revman5)  *Neurostat/3D-SSP:* sensitivity=75%; specificity=57% (Table 2, p548)  TP=31; FP=20; FN=10; TN=27 (Calculated in Revman5)  *t-sum (PET score interpretation):* sensitivity=69%; specificity=75% (Table 2, p548)  TP=28; FP=12; FN=13; TN=35 (Calculated in Revman5)  Duration of follow-up (N=88): 2 years  Visual interpretation of standard imaging: sensitivity=97%; specificity=32% (Table 2, p548)  TP=40; FP=32; FN=1; TN=15 (Calculated in Revman5)  *Neurostat/3D-SSP:* sensitivity=69%; specificity=64% (Table 2, p548)  TP=28; FP=17; FN=13; TN=30 (Calculated in Revman5)  *t-sum (PET score interpretation:* sensitivity=70%; specificity=90% (Table 2, p548)  TP=29; FP=5; FN=12; TN=42 (Calculated in Revman5)  Duration of follow-up (N=88): 3 years  Visual interpretation of standard imaging: sensitivity=98%; specificity=41% (Table 2, p548); specificity=40% (calculated in Revman5)  TP=40; FP=28; FN=1; TN=19 (Calculated in Revman5)  *Neurostat/3D-SSP:* sensitivity=64%; specificity=64% (Table 2, p548); (sensitivity=63% calculated in RevMan5)  TP=26; FP=17; FN=15; TN=30 (Calculated in Revman5)  *t-sum (PET score interpretation:* sensitivity=61%; specificity=91% (Table 2, p548)  TP=25; FP=4; FN=16; TN=43 (Calculated in Revman5)  Loss to follow-up: 23/114 participants dropped out during the follow-up period; 3/114, who converted to other dementia, were excluded from the analysis |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  The author contacted (the numbers of participants included in the analyses reported in Table 2, p548 and the accuracy of calculation in REvMan5 checked). Data for Neurostat/3D-SSP were considered in the exploratory analysis because the majority of the studies included in this exploratory analysis applied ‘computer aided visual read’; data for the longest duration of follow-up. Were included.  Authors’ conclusion  *“At 3 year follow-up, low specificity indicated that some non-converters showed AD/DLB like hypometabolism on 18F-FDG-PET images. These results were not in line with previous reports (Anchisi 2005; Arnaiz 2001; Mosconi 2004; Drzezga 2005; Silverman 2003) where higher specificity and diagnostic accuracy were reported. The true reason for low specificity in spite of a longer follow-up compared to previous reports is unclear. One possible explanation is the difference in the characteristics of registered MCI patients for each study. In fact, conversion rates from MCI to AD were very high in some studies. Such an increase in the ratio of converters may result in a decrease in false-positive cases”.*  *“Visually assessed 18F-FDG-PET is a very sensitive but relatively nonspecific measure for predicting conversion to AD in patients with MCI. On the other hand, the PET score is the most statistically significant predictive factor for conversion from MCI to AD, and the diagnostic performance of the PET score is more promising for rapid converters over 2 years”.* | | |
|  | | |
| **Lange 2016** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested case-control study; retrospective analysis of longitudinal data from ADNI database  Original sample: 108 MCI: 31 MCI-AD and 77 MCI-MCI; Validation sample: 241 MCI: 60 MCI-AD and 181 MCI-MCI; 32 healthy controls; 32 AD  We only included data on performance of the index test to discriminate between patients with MCI who convert to dementia and those who remained stable.  Exclusion criteria: conversion to non-AD dementia | |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? No  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | Original sample: 108 MCI: 31 MCI-AD and 77 MCI-MCI (downloaded from the ADNI database in March 2014). Diagnostic criteria for MCI not reported in the paper; however, it was reported elsewhere that Petersen/Winbald criteria are operationalized by ADNI (Bondi 2014).  Gender: MCI-converters: 12 women, 19 men; MCI-non-converters: 23 women, 54 men  Age (y): MCI converters 74.7±6.4 years; MCI-non-converters 74.5± 7.7  APOE ε4 carrier: not reported  MMSE: MCI converters 27.1±1.4; MCI-non-converters 27.7±1.6  Education (y): MCI converters 15.8±3.0 years; non-MCI converters 16.0±2.7  Sources of referral: not reported  Setting: multicenter  Validation sample: 241 MCI: 60 MCI-AD and 181MCI-MCI (downloaded from the ADNI database in August 2015)  Gender: MCI-converters: 23 women, 37 men; MCI-non-converters: 86 women, 95 men  Age (y): MCI converters 73.7±6.5 years; MCI-non-converters 70.5±7.2  APOE ε4 carrier: not reported  MMSE: MCI converters 27.2±1.7; MCI-non-converters 28.2±1.6  Education (y): MCI converters 16.2±2.7 years; non-MCI converters 16.3±2.6  Sources of referral: not reported  Setting: multicenter | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  In 152 out of the total of 172 subjects, FDG PET had been acquired according to a dynamic protocol so that 6 frames of 5 min duration from 30 to 60 min post injection were available for analysis. The remaining 20 FDG PETs had been acquired as 30 min static emission scan starting 30 min post injection.  *Data analysis*  Five reference points located in precuneus, left/right parietotemporal and left/right lateral temporal cortex were used. The Herzholz t-sum score approach was used. However, the PALZ tool in PMOD was not used but rather the Herholz method was implemented in a MATLAB/SPM processing pipeline. The AD-meta-ROI was created over which the t-values of the voxel-based test are summed. This AD-meta-ROI is similar to but not exactly equal to the AD-meta-ROI implemented in PMOD. Therefore, the threshold of 11090 usually used with the PALZ tool was not applicable in our analyses. The t-sum score is computed from the statistical parametric map (SPM) of the one-sided t-test of the individual patient’s FDG PET versus a database of FDG-PETs of healthy normal subjects. Thus, computation of the t-sum score requires ‘single case-SPM’. Visual reading was not perform (Dr Buchert’s email on 13/5/2016)  Threshold: not pre-specified; *ROC-analysis* was used to determine the optimal cut-off (Youden criterion) and 100 repeats of 20-fold cross-validation to avoid overly optimistic accuracy estimates; t-sum=18774 (1199), determined by the maximum Youden index  t-sum score  Original sample: MCI converters: 37817±20182; MCI-non-converters: 14400±17483  Validation sample: MCI converters: 28356±20085; MCI-non-converters: 14604±16754 | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | Target condition: conversion from MCI to Alzheimer’s disease dementia  Reference standard: ‘diagnostic trajectory’; no further information | |
| Is the reference standard likely to correctly classified the target condition? Unclear  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Unclear concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 3 years  Conversion from MCI to ADD:  *Original sample*  Number included in analysis (N=108)  At follow-up: 31 MCI-AD (disease positive); 77 MCI-MCI (disease negative)  With the SPM default setting, the t-sum score provided an **AUC=0.728**  With the SPM ‘optimized’ setting, the t-sum score provided an **AUC=0.832**  *Validation sample*  Number included in analysis (**N=241**)  At follow-up: 60 MCI-AD (disease positive); 181 MCI-MCI (disease negative)  With the SPM default setting, the t-sum score provided and **AUC=67%**; sensitivity=58%; specificity=56%; PPV=31%; NPV=80%; (Table 4, p954); TP=35; FP=80; FN=25; TN=101 (calculated in RevMan5)  With the SPM ‘optimized’ setting, the t-sum score provided and **AUC=75%** Sensitivity=70%; specificity=68%; PPV=42%; NPV=87% (Table 4, p954); TP =42; FP=58; FN=18; TN=123 (calculated in RevMan5) | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Unclear  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Unclear risk** | | |
| **Notes**  The aim of this study was to optimize the parameter settings of voxel-based SPM single subject analysis for prediction of MCI-to-AD conversion within 3 years by brain FDG PET. The following aspects of the processing pipeline were considered: frame-by-frame motion correction, [O-15]-water versus FDG template, spatial smoothing, and intensity scaling. Optimizing SPM for voxel-based single subject analysis of brain FDG PET provided considerable improvement of MCI-to-AD prediction.  Additional information were requested and obtained from the trial investigators regarding metrics used (Dr Buchert’s email on 30/5/2016) | | |
|  | | |
| **Mosconi 2004** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: prospective longitudinal cohort  Participants with aMCI, recruited over a 2-year period. Sampling procedure not described.  Exclusion criteria: major psychiatric or medical disease; using medication that could affect brain structure or function (previous subarachnoid or intra-cerebral hemorrhage, intra-cranial tumor, hydrocephalus, psychosis, major depression, alcoholism, epilepsy, ischemic stroke, vascular dementia and other dementing illnesses, anemia, untreated thyroid dysfunction, renal insufficiency, non-stabilized diabetes mellitus). | |
| Was a consecutive or random sample of patients involved? Unclear  Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 37 MCI participants diagnosed with the Petersen 2001 criteria at baseline.  Gender: converters: 5 women; 3 men; non-converters: 15 women; 14 men  Age (y): converters: 69 ± 4 years; non-converters: 63 ± 8  APOE ɛ4: total: APOE4(+)16; APOE4(-) 21. APOE4(+) MCI-non-converters: 11/16; APOE4(+) MCI-converters: 5/16; APOE4(-) MCI-non-converters: 18/21; APOE4(-) MCI-converters: 3/21  MMSE: MCI-non-converters: 28.1 ± 1.6; MCI-converters: 23.9 ± 1.7  Education (y): MCI-non-converters: 10.0 ± 5.0; MCI-converters: 8.0 ± 3.0  Sources of referral: not reported  Sources of recruitment: not reported. The recruitment was carried out according to the general protocol of the Network for Efficiency and Standardization of Dementia Diagnosis research project (Herholz 2002). | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  PET scans were performed using GE Advance PET devices (Milwaukee, WI). Scans were acquired in 2D mode with an axial field of view of 153 mm, an in-plane full width at half-maximum (FWHM) of 4.6 mm, and slice thickness of 4.25 mm. Participants were injected with a dose of 110 to 370 MBq of [¹⁸F] FDG in a resting state with eyes closed and ears unplugged in a dimly-lighted room with minimal background noise. A polycarbonate head holder was used to reduce head movement during the scan. The uptake interval between FDG injection and scan start was on average 42 ± 19 minutes. The average scan duration was 19 ± 3 minutes. Images were reconstructed using filtered back-projection including correction for attenuation measured by transmission scan and scatter using standard software as supplied by scanner manufacturers.  *Data analysis*  Basic image processing and voxel-based data analyses were performed using **SPM**99 routines (Wellcome Department of Cognitive Neurology, London, UK) implemented in MATLAB (Mathworks, Sherborn, MA). An isotropic Gaussian filter was used to smooth the spatially normalized PET images with an FWHM of 12 mm. Individual counts were normalized to mean global activity using proportional scaling to obtain relative cerebral metabolic rate for glucose (rCMRglc) values from FDG radioactivity measurements. To minimize 'edge effects' without excluding hypometabolic tissue, only those voxels with values > 80% of the mean for the whole brain were retained for all statistical analyses. Global calculation was obtained with respect to the mean voxel value.  The writers defined the precuneus (PreCu), anterior (ACC), and posterior (PCC) cingulate cortex, inferior parietal lobule (IPL), superior (STG) and middle (MiTG) temporal gyrus, and superior (SFG), middle (MiFG), and inferior frontal (IFG) gyrus, on both hemispheres, as candidate areas for possible rCMRglc alterations.  Threshold: no specific rCMRglc value is referred as threshold. PET scan was described as positive or negative for significant rCMRglc reductions in certain cerebral areas with emphasis on the inferior parietal lobule (IPL). No threshold or related quantitative data are provided.  Unclear whether the index test results interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Unclear concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standards** | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: NINCDS-ADRDA criteria  Clinicians conducting follow-up were blinded to APOE results. Unclear whether they were unaware of the ¹⁸F-FDG results. | |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 1 year (mean 12.1 ± 0.6 months)  At baseline 37 MCI.  At follow-up: 37 participants: 8 MCI-ADD; 29 MCI-MCI (p 2335)  Sensitivity: 38%; Specificity: 97% (p 2336)  Number included in analysis: 37  TP = 3; FP = 1; FN = 5; TN = 28 (calculated in Review Manager 5)  Loss to follow-up: none  All participants appear to have been included in the analyses (conversion/non-conversion outcomes were reported for 37 participants). | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes**  Additional information were requested from the trial investigators regarding the threshold but no further information was available at the time this review was prepared (email on 5th September 2013) | | |
|  |  | |
| **Nobili 2008** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested case-control study; prospective analysis of longitudinal data  36 participants with memory complaints in whom an objective memory deficit was demonstrated by means of neuropsychological tests and 17 healthy volunteers who gave their informed consent were recruited during university courses, dedicated to elderly people. Sampling procedure not described.  We only include data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable.  Exclusion criteria: presence of analphabetism, major vision disturbances, psychiatric illnesses, epilepsy, major head trauma, Parkinsonism, previous stroke or TIA and brain masses; people scoring higher than 0 on the delusion and the hallucination NPI items were excluded. | |
| Was a consecutive or random sample of patients involved? Unclear  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 36 participants with MCI, diagnosed with the Petersen 2004 criteria at baseline, were recruited from the Outpatient clinic. Demographic characteristics are reported for 33 participants who were included in the analysis.  Gender: converters: 11 women, 11 men; non-converters: 9 women, 2 men  Age (y): converters: 77.3 ± 4.8 years; non-converters: 74.6 ± 5.4  APOE ɛ4: not reported on all MCI participants; converters: 4/8 (50%); non-converters: 5/14 (36%)  MMSE: converters: 27.6 ± 1.4; non-converters: 27.4 ± 2.0  Education (y): converters: 8.5 ± 3.9; non-converters: 8.8 ± 4.7  Sources of referral: not reported  Sources of recruitment: outpatients, no further information | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  The index test was performed within 3 months from the clinical–neuropsychological examination (mean 29.9 days in participants and 29.8 days in controls). Participants fasted for at least 6 hours. Before radiopharmaceutical injection, blood glucose was checked and was < 140 mg/dl in all cases. After a 10-min rest in a silent and obscured room, with eyes closed and ears unplugged, participants were injected with approximately 370 MBq of ¹⁸F-FDG PET via a venous cannula, according to the guidelines of the European Association of Nuclear Medicine (Bartenstein 2002). They remained in the room for 30 mins after injection, and were then moved to the PET room where scanning started approximately 45 mins after injection and lasted 20 mins.  *Data analysis*  Computerized Brain Atlas (CBA; Applied Medical Imaging, Uppsala, Sweden) software was used in order to analyze neuroimaging data. Principal component analysis on Volume Region of Interest (VROI) was performed.  Threshold: not reported; visual interpretation and principal component analysis; 25 VROI used; these regions correspond to Brodmann areas (BA).  Index test was conducted before follow-up.  Unclear whether the index test results interpreted without knowledge of the results of the reference standard | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? Unclear  **Could the conduct or interpretation of the index test have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **Target condition and reference standard(s)** | Target condition: conversion from MCI to Alzheimer's disease dementia  Reference standard: NINCDS-ADRDA; DSM-IV. All participants received both reference standards.  Unclear whether clinicians conducting follow-up were aware of the ¹⁸F-FDG results. | |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 21 months on average (mean 21.1 ± 10.9 months; mean 20.6 ± 10.3 MCI/MCI; mean 22.2 ± 12.4 MCI/ADD)  At baseline: 36 MCI; at follow-up: 11 converters; 22 non-converters (Abstract)  Number included in analysis: 33  Conversion from MCI to Alzheimer’s disease dementia  TP = 9; FN = 2; TN = 20; FP = 2 (Table 4, p 2197).  Sensitivity: 82%; specificity: 91% (calculated in Review Manager 5)  Loss to follow-up: 3 participants excluded from the analysis: 2 no longer showed any cognitive objective deficit after 26 and 35 months, respectively, and were excluded from the study. Another participant developed fronto-temporal dementia, according to the current criteria (Knopman 2005) after 1 year and was excluded. | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  Additional information were requested and obtained from the trial investigators regarding metrics used; the authors reported that they did not use the standard visual (qualitative) inspection of PET scan | | |
|  | | |
| **Ossenkoppele 2012** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested case-control study; prospective data from a longitudinal MCI subset analyzed  At baseline 15 participants were included in each group: MCI, AD and controls. No further details of participant sampling and recruitment were reported.  We only included data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable.  Exclusion criteria: a history of major psychiatric or neurological illness (other than AD) and the use of nonsteroidal anti-inflammatory drugs. People with severe vascular events during the follow-up period, such as stroke or hemorrhage, were also excluded. | |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 15 participants diagnosed by the Petersen 1999 criteria. Data reported only on 12 MCI participants.  Gender: 9 men; 3 women  Age (y): mean 67 ± 7  APOE ε4 carrier: 8  MMSE: 27 ± 3  Education (y): median (range) = 6 (3 - 7)  Sources of referral: not reported.  Sources of recruitment: not reported. | |
| **Are there concerns that the included patients and setting do not much the review question? Unclear concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  150±17 MBq ¹⁸F-FDG was injected at baseline, and 35 mins later, a 10-min transmission scan (3 x 5-min frame) were performed.  *Data analysis*  For regional analysis SUVr of the frontal, parietal and lateral temporal cortices, and the medial temporal lobe and posterior cingulate were calculated.  Threshold: not reported; computer aided visual read / SUVr and ROI images  Unclear whether the index test interpreted without knowledge of the results of reference standard. | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? Unclear  **Could the conduct or interpretation of the index test have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Unclear concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | Target condition: conversion from MCI to Alzheimer's disease dementia and other forms of dementia  Reference standards: NINCDS-ADRDA criteria for AD (McKhann 1984); Reference standard for the clinical criteria for FTD not reported.  Unclear whether clinicians conducting follow-up were aware of the ¹⁸F-FDG results. | |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: mean interval 2.5 years (range 2 - 4 years)  At baseline: 15 MCI participants. Data reported only on 12 MCI participants: 4 FDG positive test; 8 FDG negative test (from the author).  At follow-up: 12 participants: 5 MCI-converters (4 MCI-ADD; 1 MCI-FTD); 7 MCI-non-converters MCI (7 MCI-MCI) (from the author).  Number included in analysis: 12; FDG positive=4; FDG negative=8  Conversion from MCI to ADD:  TP = 3; FP = 1; FN = 1; TN = 7; sensitivity=75%; specificity=88%  Conversion from MCI to all dementia:  TP = 3; FP = 1; FN = 2; TN = 6; sensitivity=60%; specificity=86%  Loss to follow-up: 3 MCI patients refused to participate in the follow-up study due to lack of motivation | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? Unclear risk** | | |
| **Notes**  We contacted the trial investigators who provided the relevant data tor the 2 x 2 table to be completed and confirmed there are no overlapping participants with the Ossenkoppele 2012b study (email on 25th July 2013). In addition, the authors confirmed that they used ‘computer aided visual read’ metric (Dr Ossenkoppele’s email on 18/5/2016) | | |
|  | | |
| **Ossenkoppele 2013** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested case-control study; prospective data of a MCI sample analyzed  154 participants included from the outpatient memory clinic for assessing the impact of molecular imaging on the diagnostic process. Among those participants there were 30 people with MCI. No further details of participant sampling and recruitment were reported.  We only include data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable.  Exclusion criteria: major clinical and psychiatric disorders, recent vascular events and excessive substance abuse. | |
| Was a consecutive or random sample of patients involved? Unclear  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 30 MCI participants diagnosed by the Petersen 2001 criteria at baseline.  Gender: 23 men; 7 women  Age (y): mean 64±9  APOE ε4 carrier: not reported  MMSE: 27±2  Education (y): not reported  Sources of referral: not reported  Sources of recruitment: Outpatient Memory Clinic, the VU University Medical Centre, The Netherlands. | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  An intravenous bolus of approximately 185 MBq of 18F-FDGwas administered. All participants rested for 15 minutes before and for 35 minutes after injection with eyes closed and ears unplugged in a dimly lit room with minimal background noise.  *Data analysis*:  Parametric SUVr images were extracted from the interval between 45 and 60 minutes after injection. In addition, 18F-FDG PET scans were analyzed using PMOD Alzheimer’s disease discrimination (PALZ) tool.  Threshold   1. Visual inspection of SUVr images & 2. The t-values of all abnormal voxels within a predefined AD mask are summed, yielding an AD t-sum that automatically classifies the scan into either normal or abnormal.   The reader had access to the clinical differential diagnosis (p 3) | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? No  If a threshold was used, was it pre-specified? Unclear  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | Target condition: conversion from MCI to Alzheimer's disease dementia and other forms of dementia  Reference standards: NINCDS-ADRDA criteria for ADD ([McKhann 1984](https://archie.cochrane.org/sections/documents/view?version=z1603171616270313284309727812803&format=REVMAN#REF-McKhann-1984)); Reference standard for the clinical criteria for FTD not reported.  Reference standards performed both with and without the index test results on the total sample. Unclear whether the data reported on 12 participants relate to the reference standards performed with or without the index test results. | |
| Is the reference standard likely to correctly classified the target condition? Unclear  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Unclear concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 2 years  At baseline: 30 MCI: 15 FDG positive test; 15 FDG negative test (Table 1, p 4)  At follow-up: 12 participants: 7 MCI-converters (6 MCI-ADD; 1 FTD); 5 MCI-non-converters MCI (5 MCI-MCI) (from the author)  Number included in analysis: 12  Conversion from MCI to ADD:  TP = 5; FP = 0; FN = 1; TN = 6; sensitivity=83%; specificity=100%  Conversion from MCI to all dementia:  TP = 5; FP = 0; FN = 2; TN = 5; sensitivity=71%; specificity=100%; FDG positive=5; FDG negative=7  Loss to follow-up:18 MCI participants. No further information. | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Unclear  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  We contacted the trial investigators and obtained relevant data tor the 2 x 2 table to be completed; it was confirmed that there are no overlapping participants with the Ossenkoppele 2012 study (email on 25th July 2013). In addition, the authors confirmed that they used both ‘computer aided visual read’ and ‘’t-sum’ metrics (Dr. Ossenkoppele’s email on 18/5/2016). | | |
|  | | |
| **Pagani 2017** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested case-control study; retrospective analysis of the longitudinal data  95 MCI-AD (converters), 27 MCI-MCI (non-converters) and 42 normal elderly participants were recruited.  We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable.  Exclusion criteria: A 15-item Geriatric Depression Scale (GDS) score of ≤10 was required for inclusion, yet a depressed trait was not considered an exclusion criterion. Patients with MRI evidence of a major stroke or brain lesions and those meeting the criteria or vascular cognitive impairment were excluded. By contrast, a Wahlund score of <3 in all regions, and the presence of white matter hyperintensities, leukoaraiosis and lacunae were not exclusion criteria. All patients had a modified Hachinski ischemic score of <3. | |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 95 aMCI patients who converted to AD within 5 years of the baseline PET scan (MCI-AD) and 27 aMCI patients who had not converted to AD after a follow-up of at least 5 years after the first FDG PET scan (MCI-MCI). Participants diagnosed by the [Petersen](https://archie.cochrane.org/sections/documents/view?version=z1603171616270313284309727812803&format=REVMAN#REF-Petersen-2001) criteria at baseline.  Gender: MCI-non-converters: 15 M, 12 F; MCI-AD converters: 31 M; 64 F  Age (y): MCI-non-converters: 71.9± 6.4; MCI-converters: 75.2 ± 5.4  APOE ɛ4: not reported  MMSE: MCI-non-converters: 26.8 ±1.5; MCI-converters: 26.0 ± 1.0  Education (y): MCI-non-converters: 8.9± 3.7; MCI-converters: 10.1 ± 2.1  Sources of referral: not reported  Sources of recruitment: memory clinic | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| **18F-FDG PET**  *Acquisition*  PET Images were acquired using a Siemens Biograph 16 PET/ CT scanner. Scans were acquired in 3-D mode with an acquisition time of 15 min. Images were reconstructed using an ordered subsets expectation maximization algorithm, with 16 subsets and six iterations, and a reconstructed voxel size of  1.33 × 1.33 × 2.00 mm.  *Data analysis*  FDG PET images were pre-processed using Statistical Parametric Mapping (SPM8) stand-alone version for spatial normalization into MNI space  (Wellcome Department of Cognitive Neurology, London, UK). A stepwise selection procedure was applied to identify the sets of regions that provided the highest discrimination. FDG uptake values were calculated in 45 anatomical VOIs in each hemisphere, as defined by the Automated Anatomical Labeling Atlas [35], and analyzed using a Matlab-based script created in-house that automatically processed the mean FDG uptake value from each of the VOIs bilaterally (‘meta-VOI’ metric). The extracted values were then normalized in each subject to the average intensity of the cerebellar VOIs on the basis that the cerebellum is poorly affected by the AD pathological process.  *Threshold*  To evaluate the accuracy, sensitivity and specificity of the method, a cut-off value was chosen as the minimum the distance from the upper left corner of the ROC curve (where specificity = sensitivity = 1). | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? No  The best accuracy was found with a set of 13-meta-VOIs (Results section)  **Could the conduct or interpretation of the index test have introduced bias? High** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | Target condition: conversion from MCI to AD dementia  Reference standard: not reported | |
| Is the reference standard likely to correctly classified the target condition? Unclear  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Unclear** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 5 years; for MCI-MCI participants mean follow-up was 7.5±1.50, range=5.0-9.8 years; for MCI-AD participants mean follow-up was 1.8±1,1, range=0.4-5.0 years  At baseline: 122 aMCI, test positive=85  At follow-up: 95 MCI-AD and 27 MCI-MCI  Number included in analysis: 122  Conversion from MCI to ADD   1. A set of three meta-VOIs; sensitivity=81%, specificity=85% (Results section)   TP = 77; FP = 4; FN = 18; TN = 23 (Calculated in RevMan5)   1. Optimized set (a set of thirteen meta-VOIs); sensitivity=87%, specificity=93% (Results section)   TP = 83; FP = 2; FN = 12; TN = 25 (Calculated in RevMan5)  Loss to follow-up: 7 aMCI who converted to DLB, and 5 aMCI who converted to FTD were excluded from the analysis. | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Unclear  Were all patients included in the analysis? No (12/134 aMCI participants who converted to non-AD dementia were excluded from the analysis)  **Could the patient flow have introduced bias? Unclear** | | |
| **Notes** | | |
|  | | |
| **Pardo 2010** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested case-control study; prospective longitudinal data analyzed  19 MCI participants and 27 healthy controls underwent extensive medical and laboratory examination. The controls were recruited from the community. Sampling procedure not described. We only include data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable.  Exclusion criteria: not reported. | |
| Was a consecutive or random sample of patients involved? Unclear  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 19 MCI participants with MCI diagnosed by the Petersen 1999 criteria at baseline.  Gender: not reported  Age (y): mean 80 years; range: 54 - 83  APOE ɛ4: not reported  MMSE: not reported  Education (y): not reported  Sources of referral: not reported  Sources of recruitment: Memory loss clinic, Geriatric, Research, Education, and Clinical Center, the Minneapolis Veterans Affairs Medical Center MVAMC) in Minneapolis, USA | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  Participants received an intravenous injection of ¹⁸F-FDG at a dose of 5 mCi/70 kg, as they reclined with eyes closed and ears open in a quiet dark room. After a 30-min uptake period, they were transferred to an ECAT 953B or ECAT Exact scanner (Siemens, Knoxville, TN). Attenuation was measured. No arterial catheters were used for absolute quantitation.  *Data analysis*  Baseline PET scan analysis was performed visually independently by two blinded, experienced physicians. The readers characterized the scans as normal or abnormal (if abnormal, ADD or FTD pattern). The patterns on which the PET readers characterized the scans as ADD or FTD are described in detail in the paper (p 328, paragraph 2.3). PET scans were adjusted to a whole-brain mean activity and stereotactically normalized using computer aided visual read/Neurostat metric. Each participant’s scan was compared voxel-wise with the normative data set after age regression to generate difference images. The patterns of metabolic change were identified visually by the readers according to criteria reported widely.  Also in 13 MCI cases and 15 controls, a computerized vector machine (SVM) classifier was applied. Cluster division of SVM (MCI template) used. The cluster were defined on the basis of the average t-image between all MCI (N=19) participants contrasted with normal controls (N=27) to provide an MCI template. Each cluster has been colorized to aid in identification.  *Threshold*  Visual interpretation; presence or absence of ‘AD pattern’ or ‘FDG pattern’.  ‘AD pattern’: hypometabolism in the medial parietal cortex (posterior cingulate/retrosplenial cortex/ precuneus) and lateral parietal region  ‘FDG pattern’: hypometabolism in anterior/superior temporal cortex and mesial/lateral prefrontal cortex  Index test was conducted and interpreted before follow-up.  Two readers of the PET scan were blinded to each other's opinions; however, they examined the images using different approaches: only ‘the transverse sections’ or ‘internet image viewer’ (iiV). | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? No  **Could the conduct or interpretation of the index test have introduced bias? High risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | Target condition: conversion from MCI to Alzheimer's disease dementia and other forms of dementia (FTD and LBD).  Reference standard: not reported.  Unclear whether clinicians conducting follow-up were aware of the ¹⁸F-FDG results. | |
| Is the reference standard likely to correctly classified the target condition? Unclear  Where the reference standard results interpreted without knowledge of the results of the index test? Unclear  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Unclear concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 3 years  At follow-up: JVP characterized the baseline PET scans of the 19 MCI participants as: 6 ADD, 1 FTD, 11 HC (healthy control), 1 artefact (non-diagnostic). In summary: 7 PET (+) participants, 11 PET (-) participants, 1 non-diagnostic (for all forms of dementia); MAK characterized the baseline PET scans of the 19 MCI participants as: 10 ADD, 1 ADD/FTD, 3 FTD, 5 HC. In summary: 14 PET (+) participants, 5 PET (-) participants (for all forms of dementia).  Number included in analysis: 18 participants for JVP Note: The participant with ‘artefact’ PET scan not included; 19 participants for MAK.  1) Conversion from MCI to ADD (Table 2, p 331).  Reader1 (JVP)  At follow-up: TP = 2; FP = 4; FN = 6; TN = 6; sensitivity=25%; specificity=60%  Reader2 (MAK):  At follow-up: TP = 3; FP = 7; FN = 6; TN = 3; sensitivity=33%; specificity=30%  Note: The PET scan read as ADD/FTD by MAK was accounted as index test (-)  2) Conversion from MCI to any form of dementia (Table 2, p 331)  Reader1 (JVP)  TP = 6; FP = 1; FN = 7; TN = 4; sensitivity=46%; specificity=80%; FDG positive=7; FDG negative=11  Reader2 (MAK)  At follow-up: TP = 9; FP = 5; FN = 5; TN = 0; sensitivity=64%; specificity=100%  Note: The PET scan read as ADD/FTD by MAK was accounted as index test (+)  Loss to follow-up: none | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Unclear  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Unclear risk** | | |
| **Notes**  The authors confirmed that they used computer aided visual read /Neurostat (Email on 22/9/2016).  In the exploratory analysis we included data from the Reader 1 who provided more accurate estimates of the diagnostic accuracy of Alzheimer's disease type imaging as it is very likely that those readers are more experienced in interpreting brain PET scans. | | |
|  | | |
| **Perani 2014** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested case-control study; retrospective analysis of longitudinal data  88 subjects with the clinical follow-up and complete diagnostic agreement with the three expert neurologists were retrospectively pooled from the population database of the Neurology Centres for Cognitive Disorders: 24 AD; 9 DLB; 24 FTLD; 28 with MCI at baseline. In addition, 112 cognitively normal subjects were included: 93 from the European  Alzheimer Disease Consortium (EADC)-PET dataset (http://www.eadc.info/) and 19 that were previously acquired in the Nuclear Medicine Department of the San Raffaele Hospital.  We only included data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable.  Exclusion criteria: not reported | |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 28 participants with MCI, diagnosed with the Petersen 2009 criteria at baseline, were included in the study.  Gender: 12 women; 16 men  Age (y): 71±5.7  APOE ε4 carrier: not reported  MMSE: range=25-28  CDR score: 0.5  Education (y): not reported  Sources of referral: not reported  Setting: data from the population database of the Neurology Centres for Cognitive Disorders in the San Raffaele Hospital, Milan, Italy | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisitions*  18F-FDG PET acquisitions of the whole patient group and of 112 cognitively normal subjects were performed according to the guidelines of the European Association of Nuclear Medicine (EANM) (Morbelli 2012; Varrone 2009). All FDG PET images of the patients were acquired at the Nuclear Medicine Dept., San Raffaele Hospital (Milan, Italy), with a Discovery STE (GE Medical Systems, Milwaukee, WI) multi-ring PET tomography (PET-CT) system (time interval between injection and scan start = 45 min; scan duration = 15 min).  *Data analysis*  Each participant scan was tested for relative ‘hypometabolism’ by comparison with a large sample of FDG PET scans from a database of normal controls on a voxel-by-voxel basis; therefore, in order to obtain voxel-based statistical parametric hypometabolic maps, each FDG-PET brain image scan was pre-processed using **SPM5** (http://www.fil.ion.ucl.ac.uk/spm/software/SPM5/), running in Matlab 7.6 (MathWorks Inc., Sherborn, MA). The FDG PET hypometabolic patterns were classified as: SPM t-map negative/AD-like/DLB-like or FTLD-like.  Visual inspection: Standard FDG images. In the first instance, the FDG PET scans were classified as: normal, uncertain and abnormal; raters were then asked to report more on brain metabolism/brain region/the pattern of brain hypometabolism, etc. (p448).  Threshold: was set up at the default 0.8 value (i.e., the mean brain intensity was computed from only those voxels with intensity above 0.8 of the mean over the entire scan). Voxel-wise comparisons were made using an explicit FDG-PET mask (Ridgway 2009).  Optimal thresholding was applied (pre-specified): Clusters of decreased metabolism were considered significant when they met a significance level of p = 0.05, corrected for multiple comparisons with the family-wise error (FWE) option at the voxel level, and contained more than 100 voxels.  Note: Nine neuroimaging experts with an extensive experience in the dementia diagnosis were presented with clinical and neuropsychological information, and with standard clinical display of FDG PET images and SPM Maps. Six possible scenarios were applied. Agreement between different pairs of raters on the same information and between the same pair of raters on different types of information was tested in order to assess reliability of the rating procedure.  In the first instance, raters were blinded to the confirmed results of the reference standard. | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | Target condition: Conversion from MCI to Alzheimer’s disease and other forms of dementia  Reference standards: NINCDS-ADRDA criteria (McKhann 2011); Reference for FTLD not reported  Follow-up clinical diagnosis was made without PET results. | |
| Is the reference standard likely to correctly classified the target condition? Unclear  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Unclear concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 27.6±4.1 months  At follow-up: 7 MCI-converters, 5 MCI-AD dementia and 2 MCI-FTLD; 15 MCI-non-converters (MCI-MCI); 6 MCI reverted to normal cognition  Number included in analysis (N=28)  Conversion from MCI to ADD (Dr Perani’s email on 15/4/16)  i) FDG PET with standard assessment: TP=2; FP=5; FN=3; TN=18; sensitivity=40%; specificity=78%; AUC=0.59  ii) FDG PET with SPM: TP=5; FP=6; FN=0; TN=17; sensitivity=100%; specificity=74%; AUC=0.87  Conversion from AD to All dementia (Dr Perani’s email on 15/4/16)  i) FDG PET with standard assessment: TP=5; FP=7; FN=2; TN=14; sensitivity=71%; specificity=67%; AUC=0.69  ii) FDG PET with SPM: TP=7; FP=11; FN=0; TN=10; sensitivity=100%; specificity=48%; AUC=0.74; FDG positive=18; FDG negative=10 | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Unclear  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Unclear risk** | | |
| **Notes**  Dr Perani (email on 09/03/2016) confirmed that there was no overlap between MCI patients included in the Perani 2014 and Perani 2015 studies. In addition, they clarified that they applied ‘computer aided visual read’ analytical approach (sc-SPM) and ‘visual read only’ (‘standard FDG images’) assessment of 18F-FDG PET.  Conclusion  *“SPM Maps showed higher sensitivity and specificity (96% and 84%), and better diagnostic positive (6.8) and negative (0.05) likelihood ratios compare to ‘Clinical Scenarios’ and ‘Standard FDG Images’.”* | | |
|  | | |
| **Perani 2016** | | |
| ***Patient selections*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | Study design: nested cross-sectional study; retrospective analysis of longitudinal data  88 patients with early dementia, including AD, frontotemporal lobar degeneration (FTLD) and dementia with Lewy bodies (DLB), and 35 participants with MCI were referred to memory clinics during the period 2009 to 2012. Two neurologists, blinded to the previous clinical diagnosis, reviewed the clinical data; of 121 participants, 16 were excluded from the initial sample. The final cohort included 75 patients with dementia and 30 patients with MCI. Five initial MCI participants were not meet criteria for MCI and were excluded.  We only included data on performance of the 18F-FDG PET index test to discriminate between patients with MCI who convert to dementia and those who remained stable.  Exclusion criteria: participants without complete diagnostic agreement between the two experts and/or not fulfilling clinical criteria for neurodegenerative dementia or for MCI were excluded | |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Yes  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | 30 MCI participants diagnosed with the Petersen 2009 criteria included in the study.  Gender: 4 women and 7 men MCI converters; 11 women and 8 men MCI-non-converters  Age (y): MCI converters 69±5.5years; non-MCI converters 68±7.6  APOE ε4 carrier: not reported  MMSE: MCI converters 26±1.8; non-MCI converters 26±2.5  Education (y): not reported  Sources of referral: not reported  Setting: memory clinics, The San Raffaele Hospital, Milan, Italy | |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  FDG PET acquisitions were performed following standardized procedures (Anchisi 2005). All images were acquired with a Discovery STE (GE Medical Systems, Milwaukee, WI) multiring PET tomography (PET/CT) system (with 45 min between injection and the start of the scan and scan duration 15 min). Each patient scan was then tested for relative “hypometabolism” by comparison with a large sample of FDG PET scans from a database of normal controls on a voxel-by-voxel basis  *Data analysis*  Each patient scan was tested for relative ‘hypometabolism’ by comparison with a large sample of FDG PET scans from a database of normal controls on a voxel-by-voxel basis. The FDG PET hypometabolic patterns were classified by two nuclear physicians, blinded to clinical/ neuropsychological details, and the CSF and MRI results. Raters were asked to classify the **SPM t-map** as normal or abnormal on the basis that a normal SPM t map should not reveal any significant hypometabolic pattern at the FEW corrected threshold, either at the voxel or the cluster level. In the event of abnormal findings, raters had to decide whether the hypometabolic pattern was suggestive of a specific neurodegenerative dementia subtype according to a well-established literature. Thus each SPM t-map was classified as negative, AD-like, DLB-like or FTLD-like on the basis of criteria available elsewhere  Threshold: was set at p=0.05, with correction for multiple comparisons using family-wise error correction (FWE) at the voxel level. Only clusters containing more than 100 voxels were deemed to be significant (Perani 2014); pre-specified  Two nuclear physicians were blinded to clinical/neuropsychological data. | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Yes  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | Target condition: Conversion from MCI to Alzheimer’s disease  Reference standard: NINCDS-ADRDA criteria | |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concern** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | Duration of follow-up: 27.48±10.43 months  At follow-up: 10 MCI-converters, 8 MCI-AD dementia and 2 MCI-non-AD dementia (1 MCI-FTLD; 1 MCI-1DLB); 20 MCI-non-converters (MCI-MCI)  Number included in analysis (N=28)  Conversion from MCI to ADD  TP=7; FP=2; FN=1; TN=18 (Dr Perani’s email on 09/03/2016); sensitivity=88%; specificity=90%; PPV=78%; NPV=95% (Calculated in RevMan5); Reported in the paper: sensitivity=86%; specificity=90%  Note: 2 MCI participants who converted to FTLD and DLB were excluded from the analysis. | |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? No  **Could the patient flow have introduced bias? High risk** | | |
| **Notes**  Dr Perani (email on 09/03/2016) confirmed that there was no overlap between MCI patients included in the Perani 2014 and Perani 2015 studies. In addition, they clarified that they applied ‘computer aided visual read’ analytical approach (sc-SPM)  In the ROC analysis, both the t-Tau/Aβ42 and the p-Tau/Aβ42 ratios were informative in the prediction of conversion to AD. Though the p-Tau/Aβ42 ratio showed a higher specificity than FDG PET (96 % vs. 90 %), FDG PET had a higher sensitivity (86 % vs. 57 %). The lower FDG PET specificity was due to the number of MCI subjects with and AD-like FDG PET pattern who did not convert during the follow-up time, and thus were considered  in the ROC analysis as false-positives  Conclusion: An Alzheimer’s disease-positive metabolic pattern as shown by FDG PET SPM in MCI was the best predictor of conversion to Alzheimer’s disease. Future multicenter studies, including larger samples of dementia patients and MCI subjects, will increase the generalizability of the present findings. The establishment of evidence-based standardized operation procedures is necessary for the correct use of biomarkers in clinical practice. | | |
|  | | |
| **Prestia 2015** | | |
| ***Patient selection*** | | |
| **A. Risk of Bias** | | |
| **Patient sampling** | | Study design: nested case-control study; retrospective analysis of longitudinal data  Patients with MCI were selected from those coming to observation at 3 independent European memory clinics: 31 from TOMC, 25 from VUMC, and 17 from KUHH  Exclusion criteria: not reported |
| Was a consecutive or random sample of patients involved? No  \*Was a case-control design avoided? Yes  Did the study avoid inappropriate exclusion? Unclear  **Could the selection of patients have introduced bias? High risk** | | |
| **B. Concerns regarding applicability** | | |
| **Patient characteristics and settings** | | 73 MCI participants diagnosed by Petersen 1999 criteria  Gender: 18 women, 11 men pMCI; 23 women, 21 men sMCI  Age (y): 67.6±8.9 (range 52-83) pMCI; 65.3±9.4 (range 50-85) MCIs  APOE ε4 carrier: 15 (58%) MCIp; 21 (51%) sMCI  MMSE: ≥24  Sources of referral: not reported  Education (y): not reported  Setting: the Translational Outpatient Memory Clinic (TOMC) of the Scientific Institute for Research and Care of Alzheimer’s Disease, IRCCS Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy; the Alzheimer Center of the VU University Medical Center (VUMC), Amsterdam, the Netherlands; and the Memory Clinic, Department of Geriatric Medicine, Karolinska University Hospital Huddinge (KUHH), Stockholm, Sweden |
| **Are there concerns that the included patients and setting do not much the review question? Low concerns** | | |
| ***Index test*** | | |
| 18F-FDG PET scan  *Acquisition*  Not reported (Available in Supplementary Material; Sections Suppl 1.2–1.4; http://dx.doi.org/10.1016/j.jalz.2014.12.001.  *Data analysis*  Regarding the chosen 18F-FDG-PET index of AD-related hypometabolism, the AD **t-sum** score combines voxel-based FDG-PET image analysis with diagnostic information on the brain regions typically affected in AD and was chosen as a summary measure of AD-related cortical hypometabolism, whereas the smallest between left and right hippocampal volumes were normalized to intracranial volumes, automatically estimated, corrected by age computing pertinent age-corrected scores (W-scores), and retained for statistical analyses. This automated summary metric of AD-related cortical hypometabolism was recently shown to have higher positive likelihood ratio than most other FDG-PET metrics.  Threshold: t-sum > 13,481 (Dr Prestia email on 18/3/2016) | | |
| **A. Risk of bias**  Where the index test results interpreted without knowledge of the results of the reference standard? Unclear  If a threshold was used, was it pre-specified? Yes  **Could the conduct or interpretation of the index test have introduced bias? Unclear risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the index test, its conduct, or interpretation differ from the review question? Low concerns** | | |
| ***Reference standard*** | | |
| **A. Risk of Bias** | | |
| **Target condition and reference standard(s)** | | Target condition: conversion from MCI to Alzheimer’s disease dementia  Reference standard: NINCDS-ADRDA criteria  Follow-up clinical diagnosis was based without knowledge of PET results |
| Is the reference standard likely to correctly classified the target condition? Yes  Where the reference standard results interpreted without knowledge of the results of the index test? Yes  **Could the reference standard, its conduct, or its interpretation have introduced bias? Low risk** | | |
| **B. Concerns regarding applicability**  **Are there concerns that the target condition as defined by the reference standard does not much the question? Low concerns** | | |
| ***Flow and timing*** | | |
| **A. Risk of Bias** | | |
| **Flow and timing** | | Duration of follow-up: total sample: mean 28±17; pMCI 23±16.3 months (range=2-76); sMCI 31.8±16.5 months (range 12-84), average 32 (only 9/44 MCI had only 12-month follow-up)  Number included in analysis (N=73)  Conversion from MCI to ADD: at follow up 29 pMCI (disease positive) and 44 sMCI (disease negative)  Sensitivity=79%; specificity=86% (p4)  TP=23; FP=6; FN=6; TN=38; PPV=79%; NPV=86% (Calculated in Revman5)  Loss to follow-up: none |
| Was there an appropriate interval between index test and reference standard? Yes  Did all patients receive the same reference standard? Yes  Were all patients included in the analysis? Yes  **Could the patient flow have introduced bias? Low risk** | | |
| **Notes**  The authors contacted; they provided the information about threshold (Dr Prestia email on 18/3/2016) | | |

MCI, mild cognitive impairment; APOE, apolipoprotein E; MMSE, Mini-Mental State Examination; 18F-FDG PET, Fluorine-18-2-fluoro-2-deoxy-D-glucose positron emission tomography; SPM, statistical parametric map; ROC, receiver operating characteristics; rCGM, Regional cerebral glucose metabolism; NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; ADD, Alzheimer's disease dementia; TP, true positive; TN, true negative; FN, false negative; FP, false positive; RevMan5, Review Manager; ROI, region of interest; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders-IV; pMCI, progressive MCI; sMCI, stable MCI; IMI, isolated memory impairment; AAMI, age-associated memory impairment; 3D-SSP, three-dimensional stereotactic surface projection; ICD-10, International Classification of Disorders; bvFTD, behavioral variant fronto-temporal; sv-PPA, semantic variant primary progressive aphasia; CBD, corticobasal degeneration; DLB, dementia with Lewy bodies; MRI, magnetic resonance imaging; PALZ, Probability of ALZheimer; SUVr, standardized uptake value ratio; PPV, positive predictive value; NPV, negative predictive value; ADNI, Alzheimer’s Disease Neuroimaging Initiative; SVM, support vector machine; CDR, Clinical dementia rating

\* We only included data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable.

**Authors contacted**

We contacted twenty-two study authors; a majority (n=16) responded. Usable data for creating two-times-two tables were obtained for eight studies [24, 38-39, 41-42, 44, 52-53]. Missing data for completing data extraction and QUADAS-2 were obtained for seven studies [40, 45-46, 48-51].

In addition, the authors of four old studies from the original review [65-68] were contacted and confirmed they had used a semi-quantitative approach and not the standard visual (qualitative) inspection of PET scan as previously reported [19].

**QUADAS 2 assessment: a brief summary (n=25)**

For *Participant Selection* only two studies [56, 17] were considered to be at low risk of bias; a few studies [17, 40, 56, 57,] used a true consecutive sample frame; risk of bias was high in twelve [24, 38, 44, 46-50, 58, 59, 61, 66] and unclear in eleven [16, 39-40, 43, 45, 57, 60, 64-65, 67-68] studies.

For the *Index test* seven studies were graded as low risk [17, 24, 39, 44-45, 48-49] and seven as unclear risk [43, 50, 59-61, 65-66]; eleven studies [16, 38, 40, 46-47, 56-58, 64, 67-68] were considered to be at high risk of bias because the threshold used was not pre-specified and the optimal cut-off level was determined from ROC analyses; consequently, the accuracy of the ¹⁸F-FDG biomarker reported in those studies might appear to be overestimated.

The risk of bias was unclear for the *Reference Standard* domain in the 16 studies [38, 40, 43, 46-48, 56-60, 64-68] mainly due to no reporting of whether clinicians conducting follow-up assessments were aware of the initial ¹⁸F-FDG biomarker analysis results; 6/16 studies did not clearly report the reference standards for diagnosing Alzheimer’s disease dementia or other types of dementia; the risk of bias was low in the nine [16-17, 24, 39, 44-45, 49-50, 61] remaining studies.

The risk of bias was high for *Flow and Timing* domain in eight studies [39-40, 45, 49, 56, 59, 65, 67] because a number of participants were excluded from the analyses and unclear in six studies [16, 46-48, 66, 68] due to poor reporting; the risk of bias was low in eleven studies [17, 24, 38, 43-44, 51, 57-58, 60-61, 64].