



This is a repository copy of *Predicting progression to Alzheimer's disease dementia using cognitive measures*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/209329/>

Version: Published Version

---

**Article:**

Macdougall, A., Whitfield, T., Needham, K. et al. (3 more authors) (2024) Predicting progression to Alzheimer's disease dementia using cognitive measures. *International Journal of Geriatric Psychiatry*, 39 (2). e6067. ISSN 0885-6230

<https://doi.org/10.1002/gps.6067>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Predicting progression to Alzheimer's disease dementia using cognitive measures

Amy Macdougall<sup>1</sup> | Tim Whitfield<sup>2</sup> | Kelly Needham<sup>3</sup> | Jonathan M. Schott<sup>4</sup> |  
Chris Frost<sup>1</sup> | Zuzana Walker<sup>2,5</sup>

<sup>1</sup>Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

<sup>2</sup>Division of Psychiatry, University College London, London, UK

<sup>3</sup>School of Health and Related Research (ScHARR), The University of Sheffield, Sheffield, UK

<sup>4</sup>Dementia Research Centre, UCL Queen Square Institute of Neurology, London, UK

<sup>5</sup>Essex Partnership University NHS Foundation Trust, Wickford, UK

## Correspondence

Zuzana Walker, Division of Psychiatry, University College London, Sixth floor, Maple House, 149 Tottenham Court Road, London W1T 7NF, UK.

Email: [z.walker@ucl.ac.uk](mailto:z.walker@ucl.ac.uk)

## Abstract

**Objectives:** It is important to determine if cognitive measures identified as being prognostic in dementia research cohorts also have utility in memory clinics. We aimed to identify measures with the greatest power to predict future Alzheimer's disease (AD) dementia in a clinical setting where expensive biomarkers are not widely available.

**Methods:** This study utilized routine Memory Clinic data collected over 18 years. From 2214 patients assessed in the clinic, we selected 328 patients with an initial diagnosis of subjective cognitive decline or mild cognitive impairment. We compared two types of statistical model for the prediction of AD dementia. The first model included baseline cognitive test scores only, while the second model also included change scores between baseline and the first follow-up.

**Results:** Baseline scores on tests of global cognitive function (Mini-mental state examination and Cambridge Cognitive Examination-Revised), verbal episodic memory and psychomotor speed were the best predictors of conversion to AD dementia. The inclusion of cognitive change scores over 1 year of follow-up improved predictive accuracy versus baseline scores alone.

**Conclusions:** We found that the best cognitive predictors of AD dementia in a clinical setting were similar to those previously identified using research cohorts. Taking change in cognitive function into account enabled the onset of AD dementia to be predicted with greater accuracy.

## KEYWORDS

Alzheimer's disease, clinical progression, mild cognitive impairment, neuropsychological tests, subjective cognitive decline

## Key points

- Whilst biomarkers of Alzheimer's disease (AD) and neurodegeneration are advancing, these are not available in all settings
- Cognitive tests are more widely available

Amy Macdougall and Tim Whitfield made an equal contribution to the work

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. International Journal of Geriatric Psychiatry published by John Wiley & Sons Ltd.

- We evaluated the prognostic utility of cognitive tests to predict incident AD dementia in a combined memory clinic sample of patients with subjective cognitive decline (SCD) or mild cognitive impairment (MCI)
- A statistical model including change in cognition (from first to second assessment) alongside baseline scores was able to predict AD more accurately versus baseline scores alone

## 1 | INTRODUCTION

Dementia is increasingly recognized as a worldwide healthcare challenge. Whilst there is still no cure for the most common cause of dementia, Alzheimer's disease (AD), two recent phase 3 trials of anti-amyloid immunotherapies delayed clinical progression in patients with mild AD.<sup>1,2</sup> It is thus imperative to identify patients in the prodromal stages of AD dementia, who can be prioritized for further investigations with the view to initiate interventions.<sup>3</sup>

The two most intensively studied at-risk states for sporadic AD dementia are subjective cognitive decline (SCD)<sup>4</sup> and mild cognitive impairment (MCI).<sup>5</sup> Briefly, SCD is defined as perceived cognitive decline without impairment on objective cognitive testing, whereas MCI is characterized by objective cognitive impairment in the absence of substantially impaired activities of daily living (ADL).

The relative risk of all-cause dementia in SCD versus age-matched healthy controls is 2.1,<sup>6</sup> while the equivalent risk for patients with MCI is considerably larger, at 15.9.<sup>7</sup> Biomarkers are increasingly used to help predict the clinical trajectory of patients with SCD and MCI.<sup>8</sup> However, many biomarkers (e.g., positron emission tomography or cerebrospinal fluid measures of beta-amyloid) are invasive, expensive and/or have limited availability outside of more developed countries. Moreover, while there is considerable interest in blood-based biomarkers of AD, these are not yet clinically validated or widely available.<sup>9</sup> Whilst the range of neuropsychological testing available in different memory clinics varies, cognitive tests are already routinely used for the assessment of dementia, including at memory clinics with fewer resources.

Whilst the number of cognitive tests used for neuropsychological assessment varies between memory clinics, it is common for a range of measures to be administered. It would be advantageous to identify those measures with the greatest power for predicting which patients will progress to AD dementia, as this could contribute to a more efficient clinical service and reduce the burden on patients. Additionally, when individuals present with SCD or MCI, and do not go on to convert to AD or another type of dementia shortly after, it is unclear which patients should be offered further clinical follow-up.<sup>10</sup> Any additional indicators regarding who should be discharged would be beneficial for clinical decision making. In addition, the advent of disease-modifying therapies means that identifying individuals on the pathway to AD dementia is increasingly important.

This study made use of 18 years of clinical data from the Essex Memory Clinic, at which patients were assessed annually using a comprehensive battery of cognitive measures. The objectives of this study were to: identify the baseline cognitive measures with the

greatest utility for predicting progression to AD dementia at the third visit (and beyond); and to establish the additional predictive power gained by incorporating data from the first follow-up after SCD/MCI diagnosis.

## 2 | MATERIALS AND METHODS

### 2.1 | Population selection

Data from the clinical database of the Essex Memory Clinic were used, spanning a period from 2002 to 2019. Only patients with a minimum of three assessments were eligible for inclusion in this study. This was to enable change over two assessments to be used to predict progression to AD dementia (hereafter, 'AD') at future visits (and compared to the predictive accuracy using baseline measures alone). However, had the outcome been measured at visit two, the association between change scores and AD could be uninterpretable due to scores from that visit also being used for diagnosis (i.e., giving rise to dependency between the predictors and outcome).

Figure 1 displays the data selection flowchart. Patients had to fulfill diagnostic criteria for SCD<sup>4</sup> or MCI<sup>5</sup> at visit 1. Assessments made before a diagnosis of SCD or MCI were excluded, so "visit 1" refers to the first assessment at which a diagnosis of SCD or MCI was made. Dementia diagnosis (including AD) was made according to ICD-10 criteria by a consensus of three senior clinicians. As part of a routine diagnostic workup, structural magnetic resonance imaging scans were performed within 6 months of the initial visit in the majority of cases and were available at the time of consensus diagnosis. Those with a diagnosis of Parkinson's disease at any time were excluded. Patients that progressed to non-AD dementia were removed from the data at time of diagnosis.

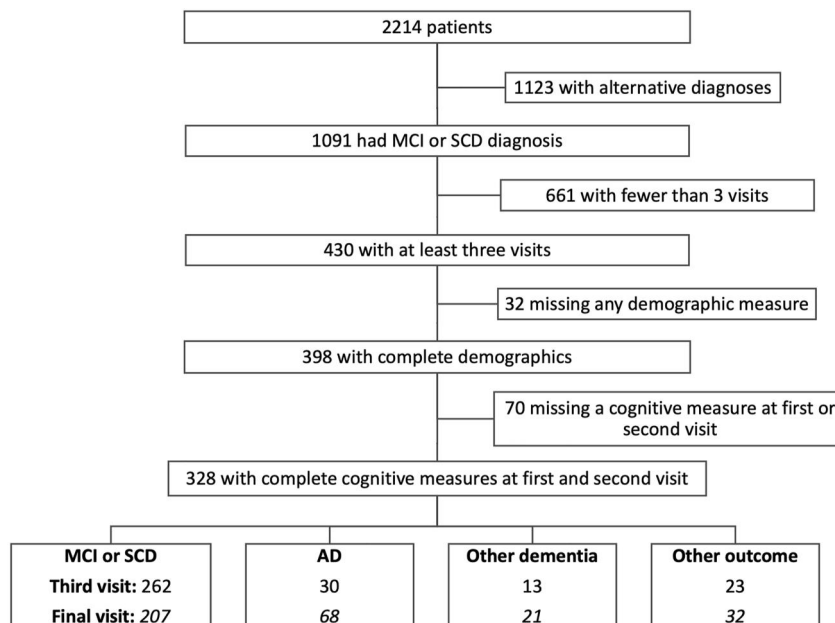
### 2.2 | Assessments

The available cognitive measures are described below:

#### 2.2.1 | Mini-mental state examination

The mini-mental state examination (MMSE) is a brief screening test for cognitive impairment.<sup>11</sup> The MMSE includes items measuring the following: orientation to time/place; registration; attention and calculation; recall; and language (maximum score 30).

**FIGURE 1** Data selection flowchart. AD, Alzheimer's Disease; MCI, Mild Cognitive Impairment; SCD, Subjective Cognitive Decline.



## 2.2.2 | Cambridge cognitive examination-revised

The Cambridge Cognitive Examination-Revised (CAMCOG-R) is a criterion test for dementia with an optimum cut-off score of 80 (maximum score 105).<sup>12</sup> Various subscales are available, but this study utilized the total score only.

## 2.2.3 | Wechsler memory scale logical memory test

The administration of the Wechsler Memory Scale Logical Memory Test (LMT) involves the assessor reading two short story paragraphs to the patient.<sup>13</sup> Immediate recall is assessed after the reading of each story, while delayed recall is assessed approximately 30 min later (scores were converted to age-adjusted percentiles for the present study).

## 2.2.4 | Trail-making test

The trail-making test (TMT) comprises two parts, each of which requires patients to link up 24 consecutive circles arranged on a page by drawing a line through them.<sup>14</sup> Part A (TMT-A) can be considered a test of simple visual attention and psychomotor speed, whereas part B (TMT-B) assesses executive task switching. For the present study, completion time in seconds was analyzed.

## 2.2.5 | Verbal fluency

Two types of verbal fluency were measured (category and letter); both measures evaluate executive control and verbal ability. For category fluency, individuals are required to generate as many animal words as

possible within 1 min.<sup>15</sup> For letter fluency, individuals are required to generate as many words as possible beginning with the letters "F", "A" and "S" (with 1 min allocated per letter).<sup>15</sup> For both tests, the total score (number of unique eligible responses) was analyzed.

## 2.2.6 | Alzheimer's disease assessment scale-cognitive (naming objects and fingers)

The Naming Objects and Fingers subtest from the AD Assessment Scale-Cognitive (ADAS-Cog) was used for the assessment of semantic memory (maximum score 17).<sup>16</sup>

Anxiety symptoms were assessed using the Rating Anxiety in Dementia (RAID)<sup>17</sup> scale or the anxiety subscale from the Hospital Anxiety and Depression Scale (HADS).<sup>18</sup> Depressive symptoms were measured using the 15-item Geriatric Depression Scale (GDS-15)<sup>19</sup> or the depression subscale from the HADS.<sup>18</sup> For patients assessed using the HADS, we used the subscales as measures of anxiety and depression, respectively. For patients who did not complete the HADS, we used the RAID score for anxiety symptoms and the GDS-15 score for depressive symptoms.

In addition, the data collected for each patient at visit one included age, sex, years of education, medical and psychiatric history, mental state examination, and physical examination (including full neurological examination).

## 2.3 | Statistical methods

### 2.3.1 | Statistical modeling

Two analyses were carried out: the first sought to predict conversion to AD at the third visit only, using logistic regression. The second

model analyzed time to conversion to AD at any visit after the second assessment using a Cox proportional hazards model.

Both analyses used a variable selection method to produce a subset of cognitive measures most strongly predictive of conversion to AD (the least absolute shrinkage and selection operator; Lasso). Cross validation was used to select the Lasso tuning parameter lambda. The value of lambda was selected which maximized the area under the receiver operating characteristic (ROC) curve for the logistic regression models, and Harrell's C-index for Cox proportional hazard models.<sup>20</sup>

Both the Lasso logistic regression and proportional hazards models were fitted in two stages. The first only included demographics and cognitive measures from the first visit. The second additionally included change in the cognitive measures between the first and second visits. These change scores were derived by subtracting visit 1 from visit 2 scores.

The Lasso method precludes the derivation of confidence intervals or *p*-values. There is also limited value in interpreting the coefficients arising from a Lasso model, as these coefficients are inherently biased. Instead, emphasis is placed on which measures were selected, rather than the size of the association between cognitive measures and the outcome. Predictive value of the selected logistic regression model was assessed using a ROC curve and the area under the curve (AUC). Harrell's C-index was used to assess the selected Cox model.<sup>20</sup>

All analyses were carried out using the statistical software R.<sup>21</sup> The package "glmnet" was used for the Lasso.<sup>22</sup>

### 2.3.2 | Predictors: Cognitive measures

All available cognitive measures were included as candidate predictors in all models. For analyses, all visit 1 cognitive measures were scaled to have a mean of zero and standard deviation (SD) of one. All change scores were also scaled using the SD of the scores from visit 1. For ease of interpretation, unscaled cognitive measures were displayed in descriptive tables. Median (interquartile range) was shown for all visit 1 scores as there was evidence of skew for most; mean (SD) was shown for change scores since these were not skewed.

We report completion times (in seconds) for TMT-A. Instead of including completion times for TMT-B, we report the ratio of TMT-A and -B completion times. This approach adjusts individuals' scores on TMT-B for their motor and visual scanning speed (indexed by TMT-A). It has been suggested that the ratio of TMT scores is a purer measure of executive functions than TMT-B alone.<sup>23</sup> In this study, the ratio was derived by dividing the completion time for TMT-A by the time for TMT-B. Thus, in-keeping with the other cognitive measures, lower scores indicate worse cognitive function (because these represent a relative increase in the time taken to complete TMT-B vs. -A). For patients who were discontinued on TMT-B by the clinician (due to making  $\geq 2$  errors), the ratio was set to zero.

### 2.3.3 | Predictors: Demographics

A limited set of demographics was selected to include in the models due to the modest sample size: age at first visit, year of first visit and presence of family history of AD (defined as a first or second degree relative with a diagnosis of AD, coded as a binary variable). Year of first visit was included to control for possible secular trends given the relatively long period of study.

## 3 | RESULTS

### 3.1 | Population

Table 1 summarizes the demographics of the patients included in this study. A total of 68 (21%) out of 328 individuals were diagnosed with AD at any point during follow-up after the second visit, 30 (44% of 68) of these were at the third visit. The median number of assessments (including visit 1) was 4. The majority of people who were diagnosed with AD during follow-up initially presented with MCI, not SCD (97% of those who progressed to AD had MCI at baseline).

The unstandardized visit 1 (baseline) and change scores between first and second visit are shown in Table 2. The average time interval between successive visits was approximately one and a quarter years. The mean visit 1 scores are worse overall for those who were (vs. were not) diagnosed with AD during follow-up. The greatest standardized mean differences are for the TMT-A, LMT delayed and MMSE, all of which exhibited a difference of  $\geq 0.4$  SD between those who do and do not convert to AD.

The change scores for the never-AD group are either close to zero or indicative of improvement for most of the cognitive measures. This most likely reflects neuropsychological retest (practice) effects.<sup>24</sup> In contrast, unstandardized change scores for those who converted to AD were mostly indicative of deterioration. This pattern of results was observed for CAMCOG-R ( $-1.10$  for the AD group,  $0.28$  for the non-AD group); LMT immediate ( $-0.24$  AD,  $6.17$  non-AD) and delayed ( $-2.26$  AD,  $7.39$  non-AD); category fluency ( $-0.40$  AD,  $0.10$  non-AD); and TMT-A/B ratio ( $-0.10$  AD,  $-0.03$  non-AD, indicating greater lengthening of the relative difference between TMT-A and -B completion times in the AD group). Again, there was a standardized mean difference of over 0.4 for the LMT delayed.

Anxiety and depression appeared to improve for both groups between visits one and two. The reduction in scores was greatest for the AD group, with mean change scores of  $-0.10$  for anxiety and  $-0.20$  for depression (compared to  $-0.07$  and  $-0.09$  in the non-AD group).

### 3.2 | Statistical models

The results of the Lasso logistic regression of diagnosis at the third visit and Lasso Cox proportional hazards regression of time to diagnosis of AD are shown in Table 3. For both analyses, "model 1"

**TABLE 1** Summary of patients from the Essex Memory Clinic included at visit 1.

	Overall	Never AD	Ever AD
Patient characteristics			
<i>n</i>	328	260	68
MCI <i>n</i> (%)	265 (80.8)	199 (76.5)	66 (97.1)
SCD <i>n</i> (%)	63 (19.2)	61 (23.5)	2 (2.9)
Gender = male <i>n</i> (%)	176 (53.7)	139 (53.5)	37 (54.4)
Year of birth (%)			
[1900–1930]	67 (20.4)	40 (15.4)	27 (39.7)
[1931–1940]	107 (32.6)	77 (29.6)	30 (44.1)
[1941–1950]	93 (28.4)	86 (33.1)	7 (10.3)
[1951–1960]	45 (13.7)	42 (16.2)	3 (4.4)
[1961–1980]	16 (4.9)	15 (5.8)	1 (1.5)
Age at first visit, years (%)			
[31–50]	20 (6.1)	19 (7.3)	1 (1.5)
[51–70]	149 (45.4)	132 (50.8)	17 (25.0)
[71–90]	159 (48.5)	109 (41.9)	50 (73.5)
Family history of AD <i>n</i> (%)	91 (27.7)	67 (25.8)	24 (35.3)
Family history of PD <i>n</i> (%)	16 (4.9)	12 (4.7)	4 (6.0)
Past psychiatric history <i>n</i> (%)	124 (37.9)	102 (39.4)	22 (32.4)
Smoking <i>n</i> (%)			
Ex-smoker	111 (33.9)	85 (32.8)	26 (38.2)
Never	173 (52.9)	139 (53.7)	34 (50.0)
Current	43 (13.1)	35 (13.5)	8 (11.8)
IQ score (mean ± SD)	0.02 (1.04)	−0.04 (1.02)	0.23 (1.09)
Years of education (median [IQR])	10 (10, 12)	10 (10, 12)	11 (9.8, 12)
Number of assessments (median [IQR])	4 [3–5]	4 [3–5]	4 [3–5]

Abbreviations: AD, Alzheimer's Disease; IQR, Interquartile Range.; MCI, Mild Cognitive Impairment; PD, Parkinson's Disease; SCD, Subjective Cognitive Decline; SD, Standard Deviation.

included baseline demographics (they were age at first visit, year of first visit, family history of AD) and measures from the first visit, while “model 2” additionally included change scores between visits one and two. A brief note on terminology: the statistical measures derived from these approaches differ, in that logistic regression yields *odds* ratios, whilst Cox models yield *hazard* ratios. The key difference between these measures of association is that hazard ratios take the time “at risk” of AD into account, while odds ratios do not (NB. though in this study, odds ratios reflected the diagnosis of AD at a relatively fixed timepoint, that is, visit 3).

Considering the results from the models which only included demographics and visit 1 scores, lower MMSE and delayed LMT scores and longer TMT-A time were associated with higher odds or hazard of AD diagnosis. In addition, in the Cox model (only), lower CAMCOG-R score was associated with greater hazard of AD.

Moving on to the results from the models which also included change scores, worsening CAMCOG-R and LMT (delayed recall)

scores were independently associated with both the odds of AD at visit 3 and hazard of AD diagnosis from visit 3 onwards.

Using observations from two visits increased the predictive power of the models: the AUC for the logistic regression model only including visit 1 scores was 0.81, compared to 0.87 for the model including two visits. The predictive power of the Cox model for time to AD diagnosis was also improved by the addition of change scores. The C-index for model 1 was 0.76, and 0.81 for model 2.

The ROC curves for the Lasso logistic regressions are shown in Figure 2. These indicate that there is a greater difference between the two models around a false positive rate (FPR) of 0.2, and less of a difference below a FPR of 0.1 or above 0.3. The model including a change score could correctly identify approximately 80% of cases of AD at the third visit, while misclassifying approximately 20% of the non-AD cases. The model including visit one only could correctly identify approximately 50%–60% of cases for a similar FPR. If the aim was to correctly identify a larger proportion of cases, over 90%, then

TABLE 2 Visit 1 and visit 1 to 2 change scores, unstandardized for ease of interpretation.

<i>n</i>	Overall 328	Never AD 260	Ever AD 68	Standardized mean difference
Time between visits, years	1.1 [1.0–1.3]	1.1 [1.0–1.3]	1.0 [1.0–1.2]	
Visit 1 (baseline) scores				
CAMCOG-R (total)	92.0 [87.0–95.0]	92.0 [87.0–96.0]	89.5 [85.0–93.2]	0.33
MMSE (total)	28.0 [26.0–29.0]	28.0 [26.0–29.0]	26.0 [25.0–28.0]	0.41
Naming test (score)	16.0 [15.0–17.0]	16.0 [15.0–17.0]	16.0 [15.0–17.0]	0.22
LMT immediate (%ile)	25.0 [7.2–63.0]	25.0 [9.0–63.0]	17.5 [5.0–50.0]	0.19
LMT delayed (%ile)	16.0 [5.0–63.0]	25.0 [5.0–63.0]	10.5 [2.0–50.0]	0.41
Letter fluency (FAS; total)	33.0 [26.0–42.0]	32.0 [26.0–42.2]	33.5 [25.8–40.5]	0.02
Category fluency (animals; total)	16.0 [13.0–20.0]	17.0 [13.0–21.0]	15.0 [12.8–18.0]	0.30
TMT-A (seconds) <sup>a</sup>	41.0 [33.0–54.0]	40.0 [31.8–52.0]	46.5 [40.0–60.0]	0.47
TMT-A/B ratio (seconds)	0.4 [0.3–0.6]	0.5 [0.3–0.6]	0.4 [0.2–0.5]	0.00
Anxiety score <sup>a,b</sup>	–0.2 [–0.8–0.4]	–0.2 [–0.7–0.4]	–0.4 [–0.9–0.3]	0.18
Depression score <sup>a,b</sup>	–0.3 [–0.8–0.4]	–0.3 [–0.8–0.7]	–0.3 [–0.8–0.2]	0.19
Visit 2–visit 1 (change) scores				
CAMCOG-R	0.0 (4.5)	0.3 (4.3)	–1.1 (5.1)	0.30
MMSE	–0.2 (2.2)	–0.2 (2.3)	–0.1 (2.1)	0.01
Naming test	0.0 (1.4)	0.0 (1.2)	0.0 (1.8)	0.01
LMT immediate	4.8 (21.8)	6.2 (21.6)	–0.2 (22.2)	0.29
LMT delayed	5.4 (23.8)	7.4 (23.8)	–2.3 (22.3)	0.42
Letter fluency (FAS)	0.8 (8.9)	0.9 (8.8)	0.5 (9.0)	0.05
Category fluency (animals)	0.0 (4.6)	0.1 (4.7)	–0.4 (3.9)	0.12
TMT-A <sup>a</sup>	1.6 (16.6)	1.8 (15.1)	0.9 (21.5)	0.05
TMT-A/B ratio	0.0 (0.5)	0.0 (0.5)	–0.1 (0.7)	0.13
Anxiety score <sup>a,b</sup>	–0.1 (0.8)	–0.1 (0.8)	–0.1 (0.7)	0.04
Depression score <sup>a,b</sup>	–0.1 (0.7)	–0.1 (0.8)	–0.2 (0.6)	0.16

Note: Visit 1 scores are median (inter-quartile range); change scores are mean (standard deviation).

Abbreviations: AD, Alzheimer's Disease; CAMCOG-R, Cambridge Cognitive Examination-Revised; LMT, Wechsler Memory Scale Logical Memory Test; MMSE, Mini-Mental State Examination; TMT, Trail-Making Test; %ile, Percentile.

<sup>a</sup>Measures for which a higher score indicates worse cognition or greater anxiety or depression; for all other measures a higher score indicates better cognition.

<sup>b</sup>Anxiety and depression scores were standardized, given these were each measured using two different scales.

both models would incur a high FPR of over 30%. Similarly, if the aim was to minimize the FPR, to under 10% for example, both models would only identify around 30% of cases.

In unplanned sensitivity analyses, patients with a history of stroke ( $n = 23$ ) were excluded, and the analyses repeated, to establish whether the same results would be obtained for a more homogeneous sample. Excluding these patients resulted in one less patient progressing to AD. The re-analysis yielded identical results for the logistic regression models (see Table 3). However, in the Cox model exclusively using visit 1 scores, only a single cognitive test was selected (i.e., TMT-A). In contrast, the Cox model using visit 1 and 2

scores was identical to the main analyses, save for the omission of baseline MMSE score.

## 4 | DISCUSSION

This study set out to identify the baseline cognitive measures with the greatest utility for predicting progression from SCD/MCI to AD dementia, and to establish the additional predictive power gained by incorporating scores from the first follow-up after SCD/MCI diagnosis.

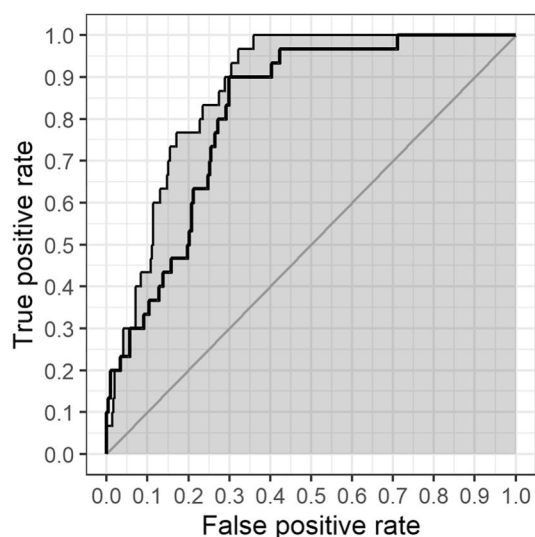
**TABLE 3** Results of Lasso logistic regressions for conversion to AD at third visit and Cox's proportional hazards models of time to conversion to AD after the second visit.

	Logistic regression with Lasso		Cox proportional hazards with Lasso	
	Model 1: Visit 1 only	Model 2: Visits 1 and 2	Model 1: Visit 1 only	Model 2: Visits 1 and 2
<b>Visit 1 scores</b>				
CAMCOG-R	-	-	0.93	0.95
MMSE	0.73	0.72	0.96	0.96
LMT delayed	0.87	0.84	0.92	0.85
TMT-A <sup>a</sup>	1.19	1.22	1.17	1.18
<b>Visit 2–visit 1 scores</b>				
CAMCOG-R	-	0.73	-	0.80
LMT delayed	-	0.79	-	0.72
Measure of predictive performance	AUC		Harrell's C-index	
	0.82	0.87	0.76	0.81

Note: Models included all cognitive measures as well as age at first visit, year of first visit, and family history of AD. Cognitive measures which were set to zero in all models by the Lasso selection method are not shown; coefficients set to zero in some models are shown as a dash.

Abbreviations: AD, Alzheimer's Disease; CAMCOG-R, Cambridge Cognitive Examination-Revised; LMT, Wechsler Memory Scale Logical Memory Test; MMSE, Mini-Mental State Examination; TMT, Trail-Making Test.

<sup>a</sup>For TMT-A, higher scores indicate worse cognition; for all other measures a higher score indicates better cognition.



**FIGURE 2** ROC curves for the Lasso logistic regression model with visit one scores (lower bold line, the area under the curve (AUC) 0.82); visit one and change scores (upper line, AUC 0.87). AUC, Area Under the Curve; ROC, Receiver Operating Characteristic.

Lower scores for the LMT (delayed recall) at baseline were consistently associated with conversion to AD at the third visit and beyond. Furthermore, deterioration in LMT delayed scores was associated with progression to AD. Given the prognostic value of LMT delayed recall in the present study, memory clinics are encouraged to consider including this measure in their cognitive batteries. Nevertheless, the LMT is a commercial product, and may

thus not be a financially viable measure in all clinical settings. Unfortunately, we are not aware of a free alternative.

We also showed that the TMT-A at the first visit was a useful predictor of future conversion to AD. The widely used CAMCOG-R and MMSE were also good predictors of conversion. CAMCOG-R was predictive both at baseline and over 1 year of follow-up. Notably, a second visit to the memory clinic after diagnosis with SCD/MCI increased the predictive power of the models as measured by the AUC and Harrell's C-index.

The finding that including change scores improved the predictive accuracy of models suggests that longitudinal cognitive data in patients with SCD/MCI captures prognostically useful information “over and above” initial scores alone. For example, having a high starting score for a given test may be less benign in individuals who subsequently show a decline; similarly, a low starting score may be less predictive if this score remains stable at follow-up. In this study, the cognitive scores of the “never AD” patients tended to improve between visit 1 and 2, whereas the “ever AD” group tended to decline. Whilst the former finding likely reflects practice effects, which can be considered a nuisance variable, a recent review found that practice effects are prognostically informative.<sup>24</sup> That is, individuals showing practice effects at follow-up had a better prognosis versus those who did not. The likely existence of practice effects in the current study may thus have contributed to the improved prediction when including change scores.

The increased predictive accuracy of an additional visit to the clinic must be weighed against the increased cost and burden on patients. The model which only included baseline scores had a reasonably good performance as measured by the AUC and C-index. As discussed in the



results section, the optimal strategy may depend on whether the aim is to: identify the most cases; keep the FPR low or strike a balance between the two. In the latter scenario, two visits appear most informative according to the data available for this study.

In sensitivity analyses, patients with SCD/MCI and a history of stroke were excluded, and the models re-estimated. Only one less patient progressed to AD versus the main analyses (likely because patients with a history of stroke were more likely to be diagnosed with vascular dementia). The inferences from the sensitivity models were largely in-keeping with the main analyses, save for the Cox model using visit 1 scores only, which only included TMT-A (perhaps reflecting that patients with prodromal AD had a different baseline cognitive profile from those with a history of stroke). This caveat notwithstanding, the results were largely robust to the inclusion/exclusion of patients with a history of stroke.

Our results were in-keeping with previous studies from research cohorts showing that tests of specific cognitive functions have utility for predicting incident AD dementia. A meta-analysis by Belleville et al.<sup>25</sup> concluded that baseline scores from verbal memory measures (e.g., LMT) and broad cognitive batteries (e.g., CAMCOG-R, MMSE) had high predictive accuracy for predicting progression from MCI to AD. It is difficult to directly compare the present results for the TMT with those of the meta-analysis. Due to an insufficient number of studies, Belleville et al. did not evaluate TMT-A, which here was a significant predictor of progression to AD. The meta-analysis did, however, find an association between poorer TMT-B performance and progression from MCI to AD.<sup>25</sup> In the present study, the ratio of TMT-A/B scores (rather than TMT-B) was evaluated as a predictor of decline, as this may be a purer measure of executive function,<sup>23</sup> but no association was observed between ratio scores and AD. Also in contrast to the present study, the meta-analysis found that tests of category fluency showed relatively high accuracy for predicting progression.<sup>25</sup> One explanation for the currently divergent findings is that, when the review authors excluded studies with a high risk of bias, the sensitivity of category fluency tests fell from 0.71 to 0.60, implying that the association between this measure and progression may not be a reliable finding. Consistent with the present results, a study of SCD prognosis using multiple cohorts found that lower MMSE score predicted progression to dementia.<sup>26</sup>

Fewer studies have investigated the utility of a *change* in cognitive scores for the prediction of future AD in patients with SCD/MCI. Of the studies that have investigated change scores, most evaluated these in combination with biomarkers. Gomar et al.<sup>27</sup> used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to predict incident AD in patients with MCI. The authors calculated 12-month change scores for cognitive measures, ADLs, and CSF/volumetric biomarkers, and selected the best combination of predictors for future AD. Change scores for TMT-B and ADLs were the only metrics to predict AD in a multivariable model.<sup>27</sup> Darmanthé et al.<sup>28</sup>

evaluated the prognostic value of plasma neurofilament light (pNfL) and MMSE to predict incident AD in ADNI patients with MCI. The addition of MMSE change scores to a model including baseline pNfL and MMSE significantly improved the prediction of AD.<sup>28</sup> Whilst, in the present study, MMSE change scores were not predictive of incident AD, scores on a more comprehensive measure of global cognition (CAMCOG-R) were.

The Lasso method was selected to choose a subset of variables most predictive of conversion to AD. There are known issues with interpretability of the coefficients produced by such models. We did not estimate the size of the associations between each cognitive measure and conversion to AD. We also did not aim to produce a clinical prediction tool. Instead, the aim was to provide a list of cognitive measures which may be most useful in clinical settings to identify patients at the highest risk of decline who might then be prioritized for follow-up. Whilst our use of the Lasso precludes direct comparisons with other studies, or direct translation of the findings to clinical practice, the measures presently found to be prognostic (or analogs thereof) are used in both research and clinical settings throughout the world. Nevertheless, in considering the implications of our findings for other memory clinics, two factors appear particularly relevant: language and cost. Whilst the MMSE and LMT are available in a wide range of languages,<sup>29,30</sup> efforts to translate the CAMCOG-R have primarily focused on European languages.<sup>31</sup> Moreover, TMT-B requires familiarity with the English alphabet sequence, putting speakers of other languages at a disadvantage. Nevertheless, culturally fair alternatives exist, including the Color Trails Test<sup>32</sup> and the Shape Trail Test.<sup>33</sup> It is also important to note that assessment materials for both the MMSE and LMT must be purchased from their respective publishers; this may not be financially viable in all settings. Whilst free and widely-translated alternatives to the MMSE exist (e.g., Addenbrooke's Cognitive Examination III,<sup>34</sup> Montreal Cognitive Assessment<sup>35</sup>), we are not aware of such alternatives to the LMT. Nevertheless, measuring cognitive changes within individual patients remains a promising strategy to improve the prediction of AD.

## 4.1 | Strengths

The longitudinal dataset used for this study was collected over 18 years from an unselected, local population. These patients were from a clinical non-academic setting, serving both rural and a wide range of urban populations (including affluent suburbia, market towns and deprived inner-city areas), thus reflecting a wide range of socio-economic circumstances.

A broad range of cognitive measures was collected at each assessment. Clinical diagnosis was made by a consensus of three senior clinicians, which is more accurate than clinical diagnosis made by a single clinician.<sup>36</sup> We undertook analyses of both conversion to AD at

the third visit, and over a longer timeframe, from the third visit onwards.

## 4.2 | Limitations

Whilst the Lasso method enabled the selection of variables most predictive of conversion to AD, the coefficients from models based on Lasso cannot be compared to one another. We were thus unable to infer which of the selected cognitive models was most predictive of AD. It was not possible with our sample to develop a classification model for clinical use; in order to do this, an even larger dataset (as well as an additional dataset for validation) would be needed.

In-keeping with clinical practice, only a limited number of patients who progressed to AD dementia had the diagnosis confirmed by amyloid positron emission tomography or brain autopsy. The data used for this study came from one clinical service in the South East of England. The results presented here should therefore be validated in datasets from other clinical settings. Lastly, two of the measures which showed utility for predicting AD (i.e., MMSE and LMT) are commercial products, which may not be financially viable in all settings. There is evidence that cognitive reserve is a protective factor against cognitive decline.<sup>37</sup> In this study, the only surrogate for cognitive reserve was the length of education, but this did not appear to have a major influence. However, this is an area that needs to be further researched.

## 4.3 | Conclusion

In this large, unselected, clinical longitudinal cohort, scores at baseline for the delayed LMT, TMT-A, CAMCOG-R and MMSE were the best predictors of conversion to AD. Notably, predictive accuracy was improved by considering change in cognitive performance over the first year of follow-up in addition to baseline scores, versus baseline scores alone. Our results from a clinical cohort were broadly in-keeping with findings from research cohorts.

## AUTHOR CONTRIBUTIONS

Substantial contributions to the conception or design of the work (Amy Macdougall, Jonathan M. Schott, Chris Frost, Zuzana Walker). The acquisition, analysis, or interpretation of data for the work (Amy Macdougall, Tim Whitfield, Kelly Needham, Jonathan M. Schott, Chris Frost, Zuzana Walker). Drafting the work or revising it critically for important intellectual content (Amy Macdougall, Tim Whitfield, Kelly Needham, Jonathan M. Schott, Chris Frost, Zuzana Walker). Final approval of the version to be published (Amy Macdougall, Tim Whitfield, Kelly Needham, Jonathan M. Schott, Chris Frost, Zuzana Walker).

## ACKNOWLEDGMENTS

The authors would like to thank the patients whose data underpinned this study. There was no specific funding in connection to this study.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

## DATA AVAILABILITY STATEMENT

The data underpinning this study were acquired during routine healthcare, and the patients did not consent to external sharing of their data. Moreover, the Health Research Authority has not approved external data sharing. As a result, unfortunately the data are not shareable.

## ETHICS STATEMENT

This project formed a part of the larger "Predictors of COgnitive DEcline in attenders of memory clinic" (CODEC) study, which was submitted to the UK Health Research Authority (HRA) for regulatory review under ID number 227614. The HRA judged that the project did not require additional review by a research ethics committee, because it exclusively included pre-existing, fully anonymized patient data. However, the project still required HRA approval, which was granted in January 2018.

## PATIENT CONSENT

All patients provided written consent for their data to be used for the purposes of this research.

## REFERENCES

- van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's disease. *N Engl J Med*. 2023;388(1):9-21. <https://doi.org/10.1056/NEJMoa2212948>
- Sims JR, Zimmer JA, Evans CD, et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA*. 2023;330(6):512-527. <https://doi.org/10.1001/jama.2023.13239>
- Long JM, Holtzman DM. Alzheimer disease: an update on pathobiology and treatment strategies. *Cell*. 2019;179(2):312-339. <https://doi.org/10.1016/j.cell.2019.09.001>
- Jessen F, Amariglio RE, van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement*. 2014;10(6):844-852. <https://doi.org/10.1016/j.jalz.2014.01.001>
- Petersen RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. *Arch Neurol*. 2005;62(7):1160-1163. discussion 1167. <https://doi.org/10.1001/archneur.62.7.1160>
- Mitchell AJ, Beaumont H, Ferguson D, Yadegarfar M, Stubbs B. Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis. *Acta Psychiatr Scand*. 2014;130(6):439-451. <https://doi.org/10.1111/acps.12336>
- Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand*. 2009;119(4):252-265. <https://doi.org/10.1111/j.1600-0447.2008.01326.x>
- van Maurik IS, Vos SJ, Bos I, et al. Biomarker-based prognosis for people with mild cognitive impairment (ABIDE): a modelling study. *Lancet Neurol*. 2019;18(11):1034-1044. [https://doi.org/10.1016/S1474-4422\(19\)30283-2](https://doi.org/10.1016/S1474-4422(19)30283-2)
- Alawode DOT, Heslegrave AJ, Ashton NJ, et al. Transitioning from cerebrospinal fluid to blood tests to facilitate diagnosis and disease monitoring in Alzheimer's disease. *J Intern Med*. 2021;290(3):583-601. <https://doi.org/10.1111/joim.13332>

10. Fonseca JA, Ducksbury R, Rodda J, et al. Factors that predict cognitive decline in patients with subjective cognitive impairment. *Int Psychogeriatr*. 2015;27(10):1671-1677. <https://doi.org/10.1017/S1041610215000356>
11. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198. [https://doi.org/10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6)
12. Roth M, Huppert F, Mountjoy CQ, Tym E. *Camdex-R: The Cambridge Examination for Mental Disorders of the Elderly*. Rev. ed. Cambridge University Press; 1999.
13. Wechsler D. *WMS-III: Administration and Scoring Manual*. Vol xiii. Psychological Corporation; 1997:217. illustrations; 24 cm.
14. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills*. 1958;8(3):271-276. <https://doi.org/10.2466/pms.1958.8.3.271>
15. Tombaugh TN, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives Clin Neuropsychology*. 1999;14(2):167-177. [https://doi.org/10.1016/S0887-6177\(97\)00095-4](https://doi.org/10.1016/S0887-6177(97)00095-4)
16. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatr*. 1984;141(11):1356-1364. <https://doi.org/10.1176/ajp.141.11.1356>
17. Shankar KK, Walker M, Frost D, Orrell MW. The development of a valid and reliable scale for rating anxiety in dementia (RAID). *Aging Ment Health*. 1999;02/01 1999;3(1):39-49. <https://doi.org/10.1080/13607869956424>
18. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361-370. <https://doi.org/10.1111/j.1600-0447.1983.tb09716.x>
19. Yesavage JA, Sheikh JI. 9/Geriatric depression scale (GDS). *Clin Gerontol*. 1986;5(1-2):165-173. [https://doi.org/10.1300/J018v05n01\\_09](https://doi.org/10.1300/J018v05n01_09)
20. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543-2546. <https://doi.org/10.1001/jama.247.18.2543>
21. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>
22. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software*. 2010;33(1):1. <https://doi.org/10.18637/jss.v033.i01>
23. Cavaco S, Goncalves A, Pinto C, et al. Trail Making Test: regression-based norms for the Portuguese population. *Arch Clin Neuropsychol*. 2013;28(2):189-198. <https://doi.org/10.1093/arclin/acs115>
24. Jutten RJ, Grandoit E, Foldi NS, et al. Lower practice effects as a marker of cognitive performance and dementia risk: a literature review. *Alzheimers Dement (Amst)*. 2020;12(1):e12055. <https://doi.org/10.1002/dad2.12055>
25. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J. Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychol Rev*. 2017;27(4):328-353. <https://doi.org/10.1007/s11065-017-9361-5>
26. Slot RER, Sikkes SAM, Berkhof J, et al. Subjective cognitive decline and rates of incident Alzheimer's disease and non-Alzheimer's disease dementia. *Alzheimers Dement*. 2019;15(3):465-476. <https://doi.org/10.1016/j.jalz.2018.10.003>
27. Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, Davies P, Goldberg TE. Alzheimer's Disease Neuroimaging I. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. *Arch Gen Psychiatr*. 2011;68(9):961-969. <https://doi.org/10.1001/archgenpsychiatry.2011.96>
28. Darmanthe N, Tabatabaei-Jafari H, Cherbun N, Alzheimer's Disease Neuroimaging I. Combination of plasma neurofilament light chain and mini-mental state examination score predicts progression from mild cognitive impairment to Alzheimer's disease within 5 years. *J Alzheimers Dis*. 2021;82(3):951-964. <https://doi.org/10.3233/JAD-210092>
29. Steis MR, Schrauf RW. A review of translations and adaptations of the Mini-Mental State Examination in languages other than English and Spanish. *Res Gerontol Nurs*. 2009;2(3):214-224. <https://doi.org/10.3928/19404921-20090421-06>
30. Dzikon C. The wechsler memory scale (WMS-IV). In: Carducci BJ, Nave CS, Carducci BJ, eds. *The Wiley Encyclopedia of Personality and Individual Differences*. John Wiley & Sons Ltd.; 2020:529-532.
31. Verhey FR, Huppert FA, Korten EC, et al. Cross-national comparisons of the Cambridge cognitive examination-revised: the CAMCOG-R: results from the European harmonization project for instruments in dementia. *Age Ageing*. 2003;32(5):534-540. <https://doi.org/10.1093/ageing/afg060>
32. Dugbartey AT, Townes BD, Mahurin RK. Equivalence of the color trails test and trail making test in nonnative English-speakers. *Arch Clin Neuropsychol*. 2000;15(5):425-431. <https://doi.org/10.1093/arclin/15.5.425>
33. Zhao Q, Guo Q, Li F, Zhou Y, Wang B, Hong Z. The Shape Trail Test: application of a new variant of the Trail making test. *PLoS One*. 2013;8(2):e57333. <https://doi.org/10.1371/journal.pone.0057333>
34. Mirza N, Panagioti M, Waheed MW, Waheed W. Reporting of the translation and cultural adaptation procedures of the Addenbrooke's Cognitive Examination version III (ACE-III) and its predecessors: a systematic review. *BMC Med Res Methodol*. 2017;17(1):141. <https://doi.org/10.1186/s12874-017-0413-6>
35. Ng KP, Chiew HJ, Lim L, Rosa-Neto P, Kandiah N, Gauthier S. The influence of language and culture on cognitive assessment tools in the diagnosis of early cognitive impairment and dementia. *Expert Rev Neurother*. 2018;18(11):859-869. <https://doi.org/10.1080/14737175.2018.1532792>
36. Thorlacius-Ussing G, Bruun M, Gjerum L, et al. Comparing a single clinician versus a multidisciplinary consensus conference approach for dementia diagnostics. *J Alzheimers Dis*. 2021;83(2):741-751. <https://doi.org/10.3233/JAD-210278>
37. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396(10248):413-446. [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6)

**How to cite this article:** Macdougall A, Whitfield T, Needham K, Schott JM, Frost C, Walker Z. Predicting progression to Alzheimer's disease dementia using cognitive measures. *Int J Geriatr Psychiatry*. 2024;e6067. <https://doi.org/10.1002/gps.6067>