Received: 29 July 2020 | Accepted: 30 July 2020 | Published online: 13 October 2020

DOI: 10.1002/dad2.12098

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

FLAME: A computerized neuropsychological composite for trials in early dementia

Helen Brooker[1] | **Gareth Williams**[2] | **Adam Hampshire**[3] | **Anne Corbett**[1] | **Dag Aarsland**[2] | **Jeffrey Cummings**[4] | **Jose Luis Molinuevo**[5] | **Alireza Atri**[6] | **Zahinoor Ismail**[1,9] | **Byron Creese**[1] | **Tormod Fladby**[7] | **Charlotte Thim-Hansen**[8] | **Keith Wesnes**[1] | **Clive Ballard**[1]

[1] St Lukes Campus, The University of Exeter Medical School, Exeter, UK

[2] Department of Old Age Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, Maurice Wohl Clinical Neuroscience Institute, London, UK

[3] Division of Brain Sciences, & Dementia Research Institute Care Research & Technology Centre, Imperial College London, London, UK

[4] Department of Brain Health, Cleveland Clinic Lou Ruvo Center for Brain Health, University of Nevada Las Vegas, Las Vegas, Nevada, USA

[5] BarcelonaBeta Brain Research Center, Hospital Clinic, Pasqual Maragall Foundation and Alzheimer's Disease and Other Cognitive Disorders Unit, Barcelona, Spain

[6] Banner Sun Health Research Institute (Arizona) & Harvard Medical School (Massachusetts), USA

[7] University of Oslo, Norway

[8] Medical & Science, Novo Nordisk A/S, Søborg, Denmark

[9] Department of Psychiatry, Clinical Neurosciences, and Community Health Sciences, Hotchkiss Brain Institute, University of Calgary, Canada

Correspondence
Clive Ballard, The University of Exeter Medical School, St Lukes Campus, Magdalen Road, Exeter, EX1 2LU, UK.
Email: c.ballard@exeter.ac.uk

Abstract

Introduction: Sensitive neuropsychological tests are needed to improve power for clinical trials in early Alzheimer's disease (AD).

Methods: To develop a neuropsychological composite (FLAME – Factors of Longitudinal Attention, Memory and Executive Function), we assessed, 10,714 participants over the age of 50 from PROTECT with validated computerized assessments for 2 years. A factorial analysis was completed to identify the key cognitive factors in all participants, and further analyses examined sensitivity to change in people with stage 2/3 early Alzheimer's disease (AD) according to the US Food and Drug Administration (FDA) framework.

Results: The FLAME composite score (speed of attention, accuracy of attention, memory, and executive function) distinguished between normal cognition and stage 2/3 early AD at baseline, and was sensitive to cognitive and global/functional decline over 2 years, with the potential to improve power for clinical trials.

Discussion: FLAME is sensitive to change, providing a straightforward approach to reduce sample size for RCTs in early AD.

Conclusion: FLAME is a useful computerized neuropsychology composite with utility for clinical trials focusing on cognition.

KEYWORDS
Alzheimer's, clinical trials, early dementia, MCI, PROTECT, sample size

1 | BACKGROUND

There are currently more than 40 million people worldwide living with Alzheimer's disease (AD) and other dementias, increasing to more than 100 million by 2050.[1] In addition, it is estimated that at least 15% of people age 60 or above have mild cognitive impairment (MCI), and that between 8% and 15% of these will progress to dementia each year, most commonly to AD.[2]

In the last 20 years there have been only two new licensed pharmacological therapies for the treatment of AD: memantine[3] and oligomannate[4] (available only in China). There are no licensed pharmacological treatments for MCI. The last decade has seen a number of high-profile unsuccessful randomized-controlled trials (RCTs) of amyloid immunotherapy and β-site amyloid precursor protein cleaving enzyme (BACE) inhibitors.[5,6] Although there are more encouraging emerging results for aducanumab, gantenerumab, and BAN2401,[7,8] there are still a very limited number of compounds in phase 2 and phase 3 clinical trials. Improved clinical trial designs are central to the treatment development pipeline. Increasingly the consensus view is that disease-modifying treatments are likely to confer the greatest benefits in people with early AD or MCI. Although this is probably correct, it raises significant challenges. Cognitive decline is often relatively modest and variable at this stage of disease.[9] The most widely used cognitive assessments are not very sensitive to subtle cognitive deficits; consequently, they require large sample sizes to confer appropriate power for clinical trials.[10] This presents a major barrier to conducting trials, adding significantly to the time and cost.

The US Food and Drug Administration (FDA) has taken a helpful step with their recent guidance for RCTs, focusing on early preclinical and clinical stages of AD. Within the guidance, early pre-clinical AD is divided into two groups: stage 2 characterized by subtle impairments on sensitive neuropsychological assessments in the absence of functional impairment and stage 3 where there are functional impairments in addition to neuropsychological deficits. The guidelines indicate that an effective intervention would have a "persuasive effect on sensitive measures of neuropsychological performance."[11] Computerized neuropsychological assessments could potentially satisfy this expectation for neuropsychological test sensitivity better than conventional pen and paper testing. With computerized assessments, test presentation is more consistent, problems with inter-rater reliability are avoided, and millisecond reaction times can be easily captured. In addition, many of the tests are free of learning effects, thereby enabling repeat testing to reduce variability.[12-15] Test batteries such as the Cognitive Drug Research battery and CogTrack, have already been shown to be sensitive to change in trials of cholinesterase inhibitors,[16] other trials of symptomatic drugs,[17] evaluation of natural products,[18] and RCTs of cognitive training.[15]

Important studies from the Alzheimer's Disease Neuroimaging Initiative (ADNI) have examined the sensitivity to change of widely used paper and pencil measures such as the Mini Mental Status Examination (MMSE) and Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS Cog) in people with pre-clinical AD/MCI. A key paper from Grill et al.,[10] focusing on a cohort of 364 people with MCI, identified a decline in cognition corresponding to a 3.64 point (standard deviation [SD] 6.76) increase on the ADAS Cog over 2 years. Assuming the same rate of decline and the same variability, a clinical trial would require a cohort of 1046 people per treatment arm to give 80% power with 5% significance (two-sided alpha) to detect a 25% treatment effect. These findings provide an important benchmark against which to test other measures. Composite neuropsychological scores, both paper and pencil and computerized, may provide an advantage over single instruments. Evolving composites include the Preclinical Alzheimer's Cognitive Composite (PACC).[19-21] Developed to maximize sensitivity to change in people who are amyloid positive, PACC has reported encouraging preliminary validation data in a longitudinal cohort of 66 participants.[19] Other emerging scales have combined neuropsychology and functional data. Encouraging baseline data for capturing changes in cognition (CatchCog) have demonstrated good correlation between a three-factor composite (memory, executive function, instrumental activities of daily living) and both informant-reported cognitive decline and cortical atrophy.[20] A 12-month follow-up study has now been reported, showing moderate to high levels of sensitivity to change in 131 people assessed longitudinally, although only 24 of the participants had either subjective cognitive decline or MCI.[21] A further scoring approach, the AD Composite Score (ADCOMS), based on specific items of commonly used dementia assessments from four MCI trials demonstrated potentially improved sensitivity over ADAS Cog and MMSE, although the level of sensitivity was less advantageous for the

RESEARCH IN CONTEXT

- Systematic review: The authors reviewed the literature using traditional (eg, PubMed) sources and meeting abstracts/presentations. Several new neuropsychology composites or scoring systems have been developed (eg, ADCOMS, CatchCog, PACC) and preliminary validation studies indicate improved sensitivity to change compared to traditional measures. The potential for a computerized neuropsychological composite to improve sensitivity to change in cognitive function has not been systematically examined in longitudinal studies.
- Interpretation: Based on a much larger cohort than has been assessed for other evolving tools, the FLAME neuropsychology composite was evaluated in 10,000 older people (including 1699 with stage 2/3 early AD - FDA framework) followed over 2 years. The FLAME composite had excellent sensitivity to change and the potential to substantially reduce sample sizes for clinical trials and compares favorably with other promising composites.
- Future directions: Further validation in cohorts assessed with AD biomarkers will be an important next step.

group with amnestic MCI.[22] Although these approaches may lead to valuable composite measures, validation is still relatively preliminary.

The CDR System, Cogstate, and CANTAB[23-26] have led the field in the provision of computerized cognitive testing, but mainly for diagnosis (CANTAB[26]) or for RCTs of symptomatic treatments such as cholinesterase inhibitors (CDR [eg[27]]). The COGSTATE[28] Brief Battery (CBB) is being employed in the Brain Health Registry. Initial findings showed that at baseline the sub-group of participants with self-reported MCI/AD diagnoses from the 6463 subjects age 55 and older had a poorer performance on the online CBB tests.[28] A composite score from COGSTATE has also been evaluated cross-sectionally in more than 4000 people from the A4 study of people with presymptomatic familial AD, showing significant but surprisingly modest differences between people who were amyloid positive and amyloid negative and modest correlations with the PACC paper and pencil composite.[29] Longitudinal data have been collected for cognitively normal individuals for each of these platforms,[23,26,30] but learning effects have been reported with both COGSTATE and CANTAB,[30,31] which could impact sensitivity to change longitudinally. There are a couple of modest longitudinal studies with CANTAB that present longitudinal data for 18 months or longer in people with early cognitive impairment. The largest of these used CANTAB to assess a cohort including 59 people with MCI followed for 18 to 24 months as part of a study focusing on subjective cognitive decline. A number of different evaluation models were presented, highlighting significantly greater decline in people with MCI than those with subjective cognitive decline in the absence of MCI.[32]

Although there are therefore some encouraging data, the potential for practice effects with some of the test batteries is a concern, and larger longitudinal studies in people with MCI and/or early AD are needed to determine sensitivity to change in these individuals across core cognitive domains. The key question of whether computerized neuropsychological assessments could provide a more sensitive approach to the development of a clinical trial neuropsychology composite for people with early AD or MCI requires more evidence from larger longitudinal studies that include at-risk populations and assess core aspects of cognitive function.[24]

The current study examines 2 years of cognitive data derived from detailed computerized neuropsychological evaluation of more than 10,000 people over the age of 50, including 1699 with a level of cognitive impairment consistent with stage 2 or stage 3 early AD, to optimize a neuropsychological composite that is sensitive to change.

2 | METHODS

2.1 | Study design

This study used 2-year longitudinal data from the online Platform for Research Online to investigate Genetics and Cognition in Ageing (PROTECT) study of people over the age of 50 without a clinical diagnosis of dementia (http://www.protectstudy.org.uk/), launched in November 2015. PROTECT was publicized through media partnerships, and interested members of the public were able to consent and join the program through the study website. The study received ethical approval from the UK London Bridge National Research Ethics Committee (Ref: 13/LO/1578). Enrollment was completed via the study website following national publicity and signposting through partner cohorts and organizations. Participants gave electronic informed consent through an online registration process. PROTECT has over 28,000 participants, and is a 10-year cohort study that allows nested clinical trials, with the main study participants undertaking annual assessments. At the time the current study commenced there were 14,201 participants. The current article reports the 10,714 participants who enrolled in a longitudinal sub-study focusing on enhanced assessment of cognition. The participants in the current study were representative of the overall study population[33]

2.2 | Data collection

All participants completed a series of online self-report questionnaires annually. This included demographic information (date of birth, gender, highest level of education) and a self-rated assessment of global change (including function) using the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE).[34,35] Although the primary evaluation was undertaken using the self-report scale, informant data were available for 4562 participants, which showed 94% concordance in the classification of participants.

Two online neuropsychological test batteries were completed (PROTECT and CogTrack), incorporating widely used and well-validated neuropsychology tests utilized by large commercial test batteries such as the CDR system. The tests of executive function (Verbal Reasoning),[15,36,25] Attention (Simple Reaction Time, Choice Reaction Time, Digit Vigilance),[37,38,18] Working Memory (Paired Associate Learning[15,39] Self-Ordered Search,[15,39] Digit Span),[15,40] and Episodic memory (Delayed Picture Recognition)[41,18,42] have been used widely over the last 35 years. The importance here is that the tests employed on PROTECT use the previously published test paradigms.

The tests and the key outputs are described in Table 1. The online cognitive assessments take approximately 30 minutes to complete and were undertaken at baseline and annually for 2 years. Through the use of parallel forms and increased sessions the tests have minimal learning effects,[12] an important issue in cognitive testing.[13,14] The participants were asked to complete the test batteries three times over 7 days annually (leaving 24 h between each session). We have demonstrated previously that there are minimal learning effects with six repeats of the tests, and that this substantially reduces variability.[12]

3 | THRESHOLDS FOR FDA STAGE 2 AND STAGE 3 EARLY AD

The FDA guidelines describe stage 2 and stage 3 early AD as pre-dementia stages of the disease.[11] Stage 2 in the FDA guideline describes individuals with neuropsychological impairments on sensitive testing in the absence of functional impairments. The FDA paper

TABLE 1 Neuropsychological test descriptions and outcome measures contributing to composites

Task			Description
PROTECT Core Neuropsychological Tests			
Self-ordered search	Total score (Memory)	5 Minutes Average[b]	A series of boxes are present on the screen; one of the boxes will contain a diamond. The volunteer selects each box until they locate the diamond. The diamond is then placed in another box and again the volunteer must locate it, but they must be careful not to select the box in which the diamond was previously found. Higher scores are achieved through efficient location of the diamond. This task measures working memory.
Paired Associate Learning	Total score (Memory)	3 Minutes Average[b]	A series of objects appear in the cells on screen. The volunteer is instructed to remember the cell in which the object appears. When an object appears at the bottom center, the volunteer is instructed to click on the cell in which they recall seeing that object. The volunteer is given three attempts at each level. This task measures working memory and learning.
Digit Span	Total score (Memory)	3 Minutes Average[b]	Using a ratchet-style approach in which each successful trial is followed by a new sequence that is one digit longer than the last and each unsuccessful trial is followed by a new sequence that is one digit shorter than the last. This task measures aspects of attention.
Verbal reasoning tasks	Total Score (executive function) Median Speed of correct responses (executive function) Standard Deviation of correct responses (executive function) Accuracy (executive function)	3 Minutes	A sentence is displayed at the bottom of the screen whilst a square and a circle are displayed above. The volunteer needs to respond as to whether the sentence correctly or incorrectly describes the configuration of the circle and square. The task measures verbal/grammatical reasoning.[c]
CogTrack Neuropsychological Tests			
Digit Vigilance	Median Speed of correct responses (Speed of Attention) Standard Deviation of correct responses (Accuracy of Attention) Accuracy (Accuracy of Attention) False Alarms (Accuracy of Attention)	3 Minutes	A target digit from one to nine is randomly selected and constantly displayed to the right-hand side of the screen. Digits are then presented one at a time in the center of the screen. The volunteer is required to respond as quickly as possible every time a digit matches the target digit. Correct detections, the speed of the detections and responses made in error (false alarms) are recorded.
Choice Reaction Time	Median Speed of correct responses (Speed of Attention) Standard Deviation of correct responses Accuracy (Accuracy of Attention)	2 Minutes	The two possible stimuli in this task that can appear on screen. Equal amounts of each stimuli type will be displayed. The volunteer is required to respond with the correct response key as quickly as possible every time the stimuli appears on screen. The accuracy and speed of each is response is recorded.
Delayed Visual Recognition (Picture Recognition)	Median Speed of correct responses (Memory) Standard Deviation of correct responses Accuracy (Memory)	3.5 Minutes	At the start of the battery 20 pictures are presented for an equal time on screen. At the end of the battery the original pictures plus the 20 very similar distractor pictures are presented one at a time in a counterbalanced order. For each picture the volunteer has to indicate whether or not it was the precise picture shown earlier, as quickly and accurately as possible. Each picture remains on the screen until a response is made. The accuracy and speed of each response is recorded.
Simple Reaction Time	Median Speed of correct responses (Speed of Attention) Standard Deviation of correct responses	2 Minutes	The volunteer is required to respond as quickly as possible when a stimulus is presented in the center of the screen. The volunteer is informed that the stimuli will be presented one at a time and that they will remain there until a response is made. The speed of each response is recorded.

[a] This is the time for a single session administration of the test.
[b] Timings wise this task is performance driven so can be as short as 1.5 minutes.
[c] A verbal reasoning task was part of both the PROTECT and CogTrack battery with different presentations and outcome measures. The CogTrack version did not restrict time for volunteers.

does not highlight specific tests or thresholds, but implies that it is a sensitive level of impairment across more than one cognitive domain. In the current study, stage 2 was operationalized to include participants who scored >1 SD below benchmarked norms[15,43-45] on at least two of the four PROTECT neuropsychological tests of memory and executive function (Table 1). Stage 3, requiring people to have neuropsychological and subtle functional impairments, was operationalized to include participants who scored >1.5 SD below benchmarked norms[15,43] on at least two of the four PROTECT tests and had a total IQCODE score ≥64.[34] The FDA paper also discusses clinically meaningful outcomes and the subsequent development of functional impairments in people at stage 2. The prediction of functional decline in patients at stage 2 is therefore an important evaluation. As the neuropsychology definition was a measure of multi-domain impairment, as a sensitivity analysis further evaluations were also undertaken for amnestic MCI (>1.5 SD from benchmarked norms in paired associate learning and functional impairment on IQCODE) and non-amnestic (executive) MCI (>1.5 SD from benchmarked norms in verbal reasoning and functional impairment on IQCODE).

4 | DATA ANALYSIS

4.1 | Factor analysis

A principal components analysis (or PCA) together with Varimax rotation was done to identify the factor structure within the neuropsychological tests undertaken using the average baseline data from all participants. Factors with eigenvalues greater than unity and factor loadings of 0.4 and above were considered to indicate that a specific task mapped to an identified factor. The derived composites were then applied to examine longitudinal decline in people with early AD (stage 2, stage 3) as a validation cohort (described in the section below).

4.2 | Main analysis

All analyses of longitudinal data were conducted using the R statistical package Version 3.6.2 Mixed Model Repeated Measure (MMRM); analyses of variance (ANOVAs) were conducted on the change from baseline data using unstructured covariance. An IQCODE baseline total score was fitted as a fixed factor. Age, gender, education, and baseline performance were fitted as covariates. Participants performed three sessions over 7 days and average performance was calculated. The effect sizes were calculated by dividing the least squares mean differences by the square root of the residual variance from the MMRM ANOVAs. Cohen's classification of effect sizes was adopted, with d = 0.2, d = 0.5, and d = 0.8 indicating the thresholds for small, medium, and large effects, respectively.[46] All scores were normalized to 0 to 100 range. The total score was determined by calculating the mean and SD of the four sub-scores, giving a total score also out of 100. The required sample sizes for clinical trials based on this change were calculated using the Harvard calculator[47] (http://hedwig.mgh.harvard.edu/sample_size/size.html), assuming parallel design with 0.05 significance level, the change SD, power of 0.8 and 25% of mean difference. Pearson correlations were undertaken to determine the relationship between change in neuropsychological performance and change in global outcome/function over 2 years measured with the IQCODE.

5 | RESULTS

5.1 | Cohort characterization

Detailed computerized neuropsychological assessments were available for 10,714 participants at baseline, of whom complete 2-year follow-up data were available for 8965 individuals (84%). A total of 1078 people with complete follow-up data met the study criteria for stage 2 early AD at baseline, and 621 individuals met study criteria for stage 3 early AD at baseline.

The cohort characteristics are described in Table 2.

Seventy-three percent of participants were female, the mean age was 61.7 (SD 7.13), and 29% of the cohort reported a family history of dementia. The PCA identified four factors: speed of attention (factor 1), accuracy of attention (factor 2), memory (factor 3) and executive function (factor 4) (Table 3).

Each factor was normalized to a maximum score of 100. The baseline scores for people with normal cognition, stage 2 early AD, and stage 3 early AD are shown in Table 4. (FLAME: Factors of Longitudinal Attention, Memory and Executive Function).

5.2 | Concurrent validity

Participants meeting the criteria for stage 2 early AD and stage 3 early AD showed significantly greater impairment on each of the four cognitive domains compared to people with normal cognition ($P < 0.0001$ for all cognitive domains). There were also significant differences in global change and function between the groups (early AD stage 2: average total IQCODE score 50.9 [SD 2.96], average item score 3.06; early AD stage 3: IQCODE average total score 67.8 [SD 3.12], average item score 4; normal cognition: average total IQCODE score 23.6 [SD 4.05], average item score 1.31, with $P < 0.01$ for all comparisons).

5.3 | Longitudinal change over 2 years

For the combined FLAME score and for each cognitive factor, a statistically significant decline in performance over 2 years was observed by group (stage 3 early AD >stage 2 early AD >normal cognition) (Table 5).

Seventeen percent of people with stage 2 or stage 3 early AD undertook regular brain training; the sensitivity to change was unaltered in a further analyses excluding these individuals.

TABLE 2 Sample characteristics

Group	Normal Cognition N = 7286		Stage 2 Early AD N = 1081		Stage 3 Early AD N = 621	
Gender	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**
N	4473	2813	852	229	492	129
Age, mean (SD)	61.1 (7.1)	63.1 (7.6)	64.9 (7.9)	61.7 (7.2)	64.2 (8.1)	61.4 (7)
Age Range	50–93	50–91	50–91	50–88	50–92	50—87
Education Breakdown (N)	University Level Education [§] (1873)	University Level Education (1109)	University Level Education (242)	University Level Education (122)	University Level Education (148)	University Level Education (81)
	Post-Secondary [†] Education (1106)	Post-Secondary Education (748)	Post-Secondary Education (297)	Post-Secondary Education (128)	Post-Secondary Education (94)	Post-Secondary Education (73)
	Secondary Education[*] (573)	Secondary Education (712)	Secondary Education (103)	Secondary Education (48)	Secondary Education (79)	Secondary Education (47)
	Vocational Qualification [‡] (623)	Vocational Qualification (542)	Vocational Qualification (98)	Vocational Qualification (43)	Vocational Qualification (44)	Vocational Qualification (55)
Group	Normal Cognition N = 7286		Stage 2 Early AD N = 1081		Stage 3 Early AD N = 621	
Gender	Female	Male	Female	Male	Female	Male
N	4473	2813	852	229	492	129
Age mean (SD)	61.1 (7.1)	63.1 (7.6)	64.9 (7.9)	61.7 (7.2)	64.2 (8.1)	61.4 (7)
Age Range	50–93	50–91	50–91	50–88	50–92	50–87
Education Breakdown (N)	University Level Education [§] (1873)	University Level Education (1109)	University Level Education (242)	University Level Education (122)	University Level Education (148)	University Level Education (81)
	Post-Secondary [†] Education (1106)	Post-Secondary Education (748)	Post-Secondary Education (297)	Post-Secondary Education (128)	Post-Secondary Education (94)	Post-Secondary Education (73)
	Secondary Education[*] (573)	Secondary Education (712)	Secondary Education (103)	Secondary Education (48)	Secondary Education (79)	Secondary Education (47)
	Vocational Qualification [‡] (623)	Vocational Qualification (542)	Vocational Qualification (98)	Vocational Qualification (43)	Vocational Qualification (44)	Vocational Qualification (55)

5.4 | Sensitivity to change in stage 2 and stage 3 early AD

Twenty-five percent of the longitudinal change in FLAME score over 2 years was calculated. This enabled an estimated sample size to be calculated for a 25% treatment effect over 2 years independently for stage 2 and stage 3 early AD. For a trial with 80% power and requiring a significance level of 0.05. Using this method, 119 people per treatment arm would be required to achieve this level of power for a trial in people with stage 2 early AD and 132 people per arm for a trial with this level of power in people with stage 3 early AD with the FLAME composite (Table 6).

Sensitivity analyses were undertaken to examine change in people with amnestic single-domain MCI (1.5 SD below benchmarked norms on PAL with functional impairment) and non-amnestic (executive) MCI (1.5 SD below benchmarked norms on grammatical reasoning). To give 80% power to 0.05 level of significance to detect a 25% treatment effect, 205 people per arm were required for a trial in people with amnestic MCI and 187 people per arm were required for a trial in people with stage executive MCI (Tables 7 and 8).

5.5 | Correlation with functional decline

Over 2 years there was a highly significant correlation between decline in neuropsychological performance on the total FLAME composite and IQCODE score as an indicator of decline in function. There was also a significant correlation between decline in each of the four individual cognitive factors in FLAME respectively and worsening function as indicated by IQCODE (Table 9).

TABLE 3 Factor analysis of 13 outcome measures from the two neuropsychological batteries

Task measure	Factor Number			
	1	2	3	4
Simple Reaction Time Speed Median	.84*	−.5	.3	−.4
Digit Vigilance Speed	.82*	−.19	.22	−.10
Choice Reaction Time Speed Median	.73*	−.6	.11	−.5
Verbal Reasoning Accuracy	−.4	.86*	−.9	.6
Verbal Reasoning Speed	−.3	.79*	−.11	.9
Verbal Reasoning Total Score	−.12	.84*	−.15	.24
Delayed Visual Recognition Accuracy	.13	−.9	.87*	−.5
Paired Associate Learning Total Score	−.5	.4	.78*	−.8
Self-Ordered Search Total Score	−.2	.27	.68*	0
Digit Span Total Score	−.8	.11	.67*	.10
Choice Reaction Time Accuracy	.32	.3	.1	.68*
Digit Vigilance Targets Correctly Detected	−.33	.1	.4	.65*
Digit Vigilance False Positive Responses	.3	−.10	.5	−.77*

Values are multiplied by 100 and rounded to the nearest integer. Values >0.4 are flagged with an*.

TABLE 4 Baseline performance on computerized FLAME composite: factors of longitudinal assessment of attention, memory and executive function

Factor Domain	Normal Cognition (N = 7286)		Stage 2 early AD (N = 1081)		Stage 3 early AD (N = 621)	
	Mean	SD	Mean	SD	Mean	SD
Speed of Attention	97.60	19.4	92.56	14.80	90.02	17.3
Accuracy of Attention	95.73	3.7	69.83	10.56	55.98	11.62
Memory	98.20	5.14	79.01	9.1	69.41	8.6
Executive function	92.8	9.58	83.04	11.1	78.12	8.4
FLAME Composite	96.40	6.81	77.61	9.83	61.08	9.36

6 | DISCUSSION

The current article presents a longitudinal study of 1699 people with neuropsychological impairments consistent with stage 2 or stage 3 early AD as part of a study of 10,000 people over the age of 50 from PROTECT. The FLAME composite score covering four domains of cognition, incorporating speed of attention, accuracy of attention, memory, and executive function, was highly sensitive to longitudinal change and predicted global decline over 2 years in people meeting the study definition of early AD and MCI . This composite may offer significant advantages for clinical trials enrolling older adults at pre-dementia stages of the disease.

TABLE 5 Differences in 2-year longitudinal change in the progression of cognitive impairment between early dementia (stage 2) and early dementia (stage 3), respectively, and people with normal cognition

	Between-group difference Normal vs stage 2 Early AD		Between-group difference Normal vs stage 3 Early AD	
	LS Mean Difference (SD)	*P* value	LS Mean Difference (SD)	*P* value
Speed of Attention	3.28 (1.73)	0.014	4.96 (1.72)	0.039
Accuracy of Attention	12.23 (1.98)	0.000	21.84 (2.29)	0.000
Memory	11.48 (2.13)	0.000	19.23 (2.97)	0.000
Executive Function	7.23 (1.32)	0.000	10.25 (2.12)	0.000
FLAME Composite	6.41 (1.48)	0.007	8.19 (2.19)	0.009

The main goal of the current study was to determine whether a computerized neuropsychological measure was more sensitive to change in people with early AD or MCI than previously reported for traditional tests. The neuropsychology composite uses widely used and well-validated computerized neuropsychological tests. The specific components of the FLAME composite were developed using baseline data from all participants in the study, including those with normal cognition. The derived composite was then applied to examine longitudinal decline in people with a level of cognitive impairment compatible with stage 2 and stage 3 early AD, using the 2018 FDA framework,[11] as a validation cohort.

Grill et al.[10] using the ADNI data set estimated that more than 1000 participants per arm would be required to detect a 25% treatment effect with 80% power to 5% level of significance using the ADAS Cog or MMSE. Using the same methodology, the FLAME computerized neuropsychology composite gives required sample sizes per treatment arm of between 119 and 205 participants per treatment arm for clinical trials focusing on early AD. Our findings suggest that this computerized approach would benefit clinical trials by substantially enhancing sensitivity to change. The FLAME composite can be conducted in the clinic, or self-directed with participants at home. This potential flexibility will give opportunities to repeat assessments more frequently without significant additional cost, and the potential to complete all assessments remotely may be particularly valuable under the challenge of coronavirus disease 2019 (COVID-19) or future pandemics and may also enable virtual monitoring of cognitive function in the community.

There are some limitations of the study. We developed operationalized study criteria for a level of cognitive and functional impairment

TABLE 6 Change in performance over 2 years on the FLAME[a] composite and individual factors in people with stage 2 and stage 3 early AD and potential power for clinical trials

	Stage 2 Early AD (n = 1087)					Stage 3 early AD (n = 621)				
Factors	LS Mean Decline over 2 years	SD	F value	*P* value	Number per arm for 25% treatment effect (80% power, $P < 0.05$)	LS Mean Decline over 2 years	SD	F value	*P* value	Number per arm for 25% treatment effect (80% power, $P < 0.05$)
Speed of Attention	3.54	3.18	27.35	0	404	3.62	4.32	14.32	0	726
Accuracy of Attention	4.85	2.61	97.56	0	150	7.97	3.19	39.08	0	84
Memory	4.78	2.49	42.85	0	138	7.33	3.22	21.37	0	100
Executive Function	2.94	2.08	90.22	0	250	3.09	2.23	27.94	0	266
FLAME Composite	4.03	2.74	41.16	0.004	119	5.5	3.97	60.83	0.002	132

[a]Factors of longitudinal, Attention, Memory, and Executive Function.

TABLE 7 Change in neuropsychological performance on the flame composite and the individual factors over 2 years in people with amnestic MCI

Population	LS Mean Change from Baseline	SD	F Value	*P* Value	25% value of mean diff	Sample Size Needed (Parallel)
N = 548 MCI Amnestic						
Speed of Attention	−1.52	1.48	70.43	<.0001	−0.38	480
Accuracy of Attention	−3.17	2.97	4.9	0.0002	−0.79	446
Memory	−3.28	2.43	2.37	0.0093	−0.82	278
Executive Function	−1.93	2.14	5.26	0.014	−0.48	626
FLAME Composite	−2.97	2.38	21.8	0.0016	0.74	328

TABLE 8 Change in neuropsychological performance on the FLAME composite and the individual Factors over 2 years in people with executive MCI

Population	LS Mean Change from Baseline	SD	F Value	*P* Value	25% value of mean diff	Sample Size Needed (Parallel)
N = 514 Executive MCI						
Speed of Attention	−2.16	2.06	7.42	0.0057	−0.54	460
Accuracy of Attention	−3.49	3.11	12.26	<.0001	−0.87	404
Memory	−3.29	2.16	3.21	0.054	−0.82	220
Executive Function	−1.73	2.41	7.24	0.0008	−0.43	990
FLAME Composite	−2.41	2.19	16.54	0.0003	0.60	422

compatible with stage 2 and stage 3 AD using the FDA framework, but the nature of the study precluded the assessment of either Alzheimer's biomarkers (including amyloid) or cerebrovascular disease, which are acknowledged as limitations. Studies in amyloid-positive individuals will be important for further validation, although based on data using other cognitive assessments, sensitivity to change is likely to be greater in these individuals.[10,48] It should also be noted that ADAS Cog was not evaluated in the current study, and therefore this comparison is being made with published data from another cohort. It will also be important to undertake comparisons with other evolving composites such as PACC, Catch-Cog, and ADCOMS as further data emerges from other studies of these potentially promising measures. The neuropsychological tests incorporated in the composite used picture recognition as a key test of episodic memory, which is language free, rather than word recall. The task utilizes similar picture pairs (validated during the original test development), asking subjects to distinguish the original from novel. The task always presented pictures from the same 20 categories. Compared to verbal episodic memory tasks, visual memory tasks reduce bias of language and education, thus being advantageous for multi-center trials with diverse participants. It should be acknowledged though that these assessments were based on recognition and not verbal recall, as recall-based episodic memory tasks are difficult to incorporate into remote computerized tools. The breadth of domains incorporated within FLAME also make the composite potentially applicable for studies of non-AD or mixed dementias and conditions that typically produce cognitive profiles that heavily impact

TABLE 9 Correlation between change in neuropsychological performance on the FLAME composite and change in function (as indicated by IQCODE) over 2 years

		IQCODE SELF (n = 8965)	IQCODE INFORMANT (n = 4562)
Speed of Attention	Pearson Correlation	0.84	0.91
	Sig. (two-tailed)	0.001	0.043
Accuracy of Attention	Pearson Correlation	0.612	0.69
	Sig. (two-tailed)	0.003	0.011
Memory	Pearson Correlation	0.732	0.93
	Sig. (two-tailed)	0.005	0.016
Executive Function Factor	Pearson Correlation	0.589	0.732
	Sig. (two-tailed)	0.001	0.003

attention, cognitive processing speed, and executive functions, such as "subcortical" dementia, but further validation studies are needed.

7 | CONCLUSION

FLAME is a sensitive measure of change in people with stage 2 and stage 3 early dementia and MCI and provides a highly practical and straightforward approach to improving power and reducing sample size for clinical trials focusing on people with early AD, and with potential utility for virtual monitoring of cognitive deficits in the community.

FUNDING SOURCES/ACKNOWLEDGEMENTS

NIHR Maudsley Biomedical Research Centre for Mental Health

Biomedical Research Centre for mental health at South London and Maudsley NHS Foundation Trust and King's College London, Alzheimer's Society, UK.

Sponsor: The University of Exeter, UK

CONFLICTS OF INTEREST

Helen Brooker reports personal fees from Wesnes Cognition Ltd, owner of CogTrack, personal fees from University of Exeter, outside the submitted work. Professor Williams reports no conflict of interests. Professor Hampshire is owner and director of Future Cognition Ltd, which develops bespoke online cognitive tests for third parties. Dr. Corbett has nothing to disclose. Professor Aarsland reports grants and personal fees from Astra-Zeneca, grants and personal fees from H. Lundbeck, grants and personal fees from Novartis Pharmaceuticals, grants and personal fees from GE Health, grants and personal fees from Eisai, and grants and personal fees from Axovant, outside the submitted work. This paper represents independent research [part] funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. Dr. Cummings has provided consultation to Acadia, Actinogen, AgeneBio, Alkahest, Alzheon, AnnovisBio, Avanir, Axsome, Biogen, BioXcel, Cassava, Cerecin, Cerevel, Cortexyme, EIP Pharma, Eisai, Foresight, GemVax, Genentech, Green Valley, Grifols, Karuna, Merck, Novo Nordisk, Otsuka, Resverlogix, Roche, Samumed, Samus, Signant Health, Suven, Third Rock, and United Neuroscience pharmaceutical and assessment companies. Dr. Cummings has stock options in ADAMAS, AnnovisBio, MedAvante, BiOasis. Dr. Cummings owns the copyright of the Neuropsychiatric Inventory. Dr Cummings is supported by Keep Memory Alive (KMA); National Institute of Health (NIH) grant P20GM109025; National Institute of Stroke and neurological Disease grant U01NS093334; and National Institute of Aging grant R01AG053798. Jose Luis Molinuevo has served/serves as a consultant or at advisory boards for the following for-profit companies, or has given lectures in symposia sponsored by the following for-profit companies: Roche Diagnostics, Genentech, Novartis, Lundbeck, Oryzon, Biogen, Lilly, Janssen, Green Valley, MSD, Eisai, Alector, BioCross, GE Healthcare, ProMIS Neurosciences, and NovoNordisk. Alireza Atri reported receiving honoraria for consulting; participating in independent data safety monitoring boards; providing educational lectures, programs, and materials; or serving on advisory boards for Allergan, the Alzheimer's Association, Axovant, Biogen, Grifols, Harvard Medical School Graduate Continuing Education, Lundbeck, Merck, Roche/Genentech, Sunovion, and Suven; receiving book royalties from Oxford University Press; and having institutional contracts or receiving investigational clinical trial–related funding from the American College of Radiology, AbbVie, Avid, Biogen, Lilly, Lundbeck, Merck, and vTV Therapeutics. Dr. Ismail reports grants and personal fees from Janssen, personal fees from Lundbeck, and personal fees from Otsuka, outside the submitted work. Dr. Creese has nothing to disclose. Professor Fladby reports no conflicts of interest. Dr Hansen is an employee of Novo Nordisk A/S and holds stock/shares in Novo Nordisk A/SÐ outside the submitted work. Keith Wesnes is CEO of Cogtrack. Professor Ballard reports grants and personal fees from Acadia pharmaceutical company, grants and personal fees from Lundbeck, personal fees from Roche, personal fees from Otsuka, personal fees from Biogen, personal fees from Eli Lilly, personal fees from Novo Nordisk, personal fees from AARP, grants and personal fees from Synexus, and personal fees from Exciva outside the submitted work.

REFERENCES

1. World Alzheimer Report Alzheimer's Disease International 2016. https://www.alz.co.uk/research/world-report-2016
2. Petersen RC. Mild cognitive impairment. *Continuum (Minneap Minn)*. 2016;22(2 Dementia):404-418.
3. Reisberg B, Doody R, Stöffler A, Schmitt F, Ferris S, Möbius HJ. Memantine in moderate-to-severe Alzheimer's disease. *N Engl J Med*. 2003;348(14):1333-1341.
4. Syed YY. Sodium oligomannate: first approval. *Drugs*. 2020;80(4):441-444.

5. Egan MF, Kost J, Voss T. Randomized trial of verubecestat for prodromal Alzheimer's disease. *N Engl J Med*. 2019;380(15):1408-1420.
6. Honig LS, Vellas B, Woodward M, et al. Trial of solanezumab for mild dementia due to Alzheimer's disease. *N Engl J Med*. 2018;378(4):321-330.
7. https://www.marketscreener.com/BIOGEN-INC-4853/news/Biogen-Aducanumab-Update-Presentation-29432882/2019
8. Marciani DJ. Promising results from Alzheimer's disease passive immunotherapy support the development of a preventive vaccine. *Research (Wash D C)*. 2019;2019:5341375.
9. Veitch DP, Weiner MW, Aisen PS, et al. Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's disease Neuroimaging Initiative. *Alzheimers Dement*. 2019;15(1):106-152.
10. Grill JD, Di L, Lu PH, et al. Estimating sample sizes for predementia Alzheimer's trials based on the Alzheimer's Disease Neuroimaging Initiative. *Neurobiol Aging*. 2013;34(1):62-72.
11. FDA 2018 https://www.fda.gov/media/110903/download
12. Ballard C, Brooker H, Khan Z, Corbett A, Aarsland D, Wesnes K. Repeated cognitive testing reduces sample size necessary to detect clinically relevant effects. *Alzheimers Dement*. 2018;14(7):P595.
13. Wesnes K, Pincock C. Practice effects on cognitive tasks: a major problem?. *Lancet Neurol*. 2002;1(8):473.
14. Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimers Dement*. 2015;1(1):103-111.
15. Corbett A, Owen A, Hampshire A, et al. The effect of an online cognitive training package in healthy older adults: an online randomized controlled trial. *J Am Med Dir Assoc*. 2015;16(11):990-997.
16. McKeith I, Del Ser T, Spano P, et al. Efficacy of rivastigmine in dementia with Lewy bodies: a randomised, double-blind, placebo-controlled international study. *Lancet North Am Ed*. 2000;356(9247):2031-2036.
17. Wesnes KA, Aarsland D, Ballard C, Londos E. Memantine improves attention and episodic memory in Parkinson's disease dementia and dementia with Lewy bodies. *Int J Geriatr Psychiatry*. 2015;30(1):46-54.
18. Wesnes KA, Ward T, McGinty A, Petrini O. The memory enhancing effects of a Ginkgo biloba/Panax ginseng combination in healthy middle-aged volunteers. *Psychopharmacology (Berl)*. 2000;152(4):353-361.
19. Bransby L, Lim YY, Ames D, et al. Sensitivity of a Preclinical Alzheimer's Cognitive Composite (PACC) to amyloid β load in preclinical Alzheimer's disease. *J Clin Exp Neuropsychol*. 2019;41(6):591-600.
20. Jutten RJ, Harrison JE, Lee Meeuw Kjoe PR, et al. Assessing cognition and daily function in early dementia using the cognitive-functional composite: findings from the Catch-Cog study cohort. *Alzheimers Res Ther*. 2019;11(1):45.
21. Jutten RJ, Harrison JE, Brunner AJ, et al. The Cognitive-functional composite is sensitive to clinical progression in early dementia: longitudinal findings from the Catch-Cog study cohort. *Alzheimers Dement (N Y)*. 2020;6(1):e12020.
22. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. *J Neurol Neurosurg Psychiatry*. 2016;87(9):993-999.
23. Collerton J, Collerton D, Arai Y, et al. A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ Pilot Study. *J Am Geriatr Soc*. 2007;55(10):1630-1635.
24. Wesnes KA. Moving beyond the pros and cons of automating cognitive testing in pathological aging and dementia: the case for equal opportunity. *Alzheimers Res Ther*. 2014;6:58.
25. Nosheny RL, Camacho MR, Insel PS, et al. Online study partner-reported cognitive decline in the Brain Health Registry. *Alzheimers Dement (N Y)*. 2018;4:565-574.
26. Barnett JH, Blackwell AD, Sahakian BJ, Robbins TW. The Paired Associates Learning (PAL) test: 30 years of **CANTAB** translational neuroscience from laboratory to bedside in dementia research. *Curr Top Behav Neurosci*. 2016;28:449-474.
27. Wesnes KA, McKeith I, Edgar C, Emre M, Lane R. Benefits of rivastigmine on attention in dementia associated with Parkinson disease. *Neurology*. 2005;65(10):1654-1656.
28. Mackin RS, Insel PS, Truran D, et al. Unsupervised online neuropsychological test performance for individuals with mild cognitive impairment and dementia: results from the Brain Health Registry. *Alzheimers Dement (Amst)*. 2018;10:573-582.
29. Papp KV, Rentz DM, Maruff P, et al, on behalf of the A4 Study Team. The computerized cognitive composite (c3) in a4, an alzheimer's disease secondary prevention trial. *J Prev Alz Dis*. 2020. https://doi.org/10.14283/jpad.2020.38. Published online June 19, 2020.
30. Stricker NH, Lundt ES, Alden EC, et al. Longitudinal comparison of in clinic and at home administration of the cogstate brief battery and demonstrated practice effects in the Mayo Clinic Study of Aging. *J Prev Alzheimers Dis*. 2020;7(1):21-28.
31. Cacciamani F, Salvadori N, Eusebi P, et al. Evidence of practice effect in CANTAB spatial working memory test in a cohort of patients with mild cognitive impairment. *Appl Neuropsychol Adult*. 2018;25(3):237-248.
32. Campos-Magdaleno M, Leiva D, Pereiro AX, et al. Changes in visual memory in mild cognitive impairment: a longitudinal study with CANTAB. *Psychol Med*. 2020:1-11. https://doi.org/10.1017/S0033291720001142. Online ahead of print.
33. Huntley J, Corbett A, Wesnes K, et al. Online assessment of risk factors for dementia and cognitive function in healthy adults. *Int J Geriatr Psychiatry*. 2018;33(2):e286-e293.
34. Jorm AF. The informant questionnaire on cognitive decline in the elderly (IQCODE): a review. *Int Psychogeriatr*. 2004;16(3):275-293.
35. Park MH. Informant questionnaire on cognitive decline in the elderly (IQCODE) for classifying cognitive dysfunction as cognitively normal, mild cognitive impairment, and dementia. *Int Psychogeriatr*. 2017;29(9):1461-1467.
36. Baddeley AD. A three-minute reasoning test based on grammatical transformation. *Psychometric science*. 1968;10:341-342.
37. Fearn S, Hutchinson S, Riding G, Hill-Wilson G, Wesnes K, McCollum C. Carotid endarterectomy improves cognitive function in-patients with exhausted cerebrovascular reserve. *Eur J Vasc Endovasc Surg*. 2003;26:529-536.
38. Kennedy DO, Haskell CF, Wesnes KA, Scholey AB. Potentiation of cognitive effects of Ginkgo biloba by complexation with phosphatidylserine. *J Psychopharmacol*. 2004;18(Suppl):A49.
39. O'Connell H, Coen R, Kidd N, Warsi M, Chin AV, Lawlor BA. Early detection of Alzheimer's disease (AD) using the CANTAB paired Associates Learning Test. *Int J Geriatr Psychiatry*. 2004;19(12):1207-1208.
40. Owen AM, Downes JJ, Sahakian BJ, et al. Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*. 1990;28:1021e1034.
41. Conklin HM, Curtis CE, Katsanis J, Iacono WG. Verbal working memory impairment in schizophrenia patients and their first-degree relatives: evidence from the digit span task. *Am J Psychiatry*. 2000;157:275e277.
42. Wesnes KA, Brooker H, Ballard C, McCambridge L, Stenton R, Corbett A. Utility, reliability, sensitivity and validity of an online test system designed to monitor changes in cognitive function in clinical trials. *Int J Geriatr Psychiatry*. 2017;32(12):e83-e92.

43. Owen AM, Hampshire A, Grahn JA, et al. Putting brain training to the test. *Nature*. 2010;465(7299):775-778.
44. Hampshire A, Highfield RR, Parkin NBL, Owen AM. Fractionating human intelligence. *Neuron*. 2012;76(6):1225-1237.
45. Jolly AE, Scott GT, Sharp DJ, Hampshire AH. Distinct patterns of structural damage underlie working memory and reasoning deficits after traumatic brain injury. *Brain*. 2020;143(4):1158-1176.
46. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic; 1988.
47. (http://hedwig.mgh.harvard.edu/sample_size/size.html),
48. Ballard C, Atri A, Boneva N, et al. Enrichment factors for clinical trials in mild-to-moderate Alzheimer's disease. *Alzheimers Dement (N Y)*. 2019;5:164-174.

How to cite this article: Brooker H, Williams G, Hampshire A, et al. FLAME: A computerized neuropsychological composite for trials in early dementia. *Alzheimer's Dement*. 2020;12:e12098. https://doi.org/10.1002/dad2.12098