doi:10.1093/brain/awp041 Brain 2009: 132; 1299–1309 | 1299

BRAIN
A JOURNAL OF NEUROLOGY

A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia

Teresa Torralva,[1,2] María Roca,[1,2] Ezequiel Gleichgerrcht,[1] Tristán Bekinschtein[1,*] and Facundo Manes[1,2]

1 Institute of Cognitive Neurology (INECO), Buenos Aires, Argentina
2 Institute of Neurosciences Favaloro University, Buenos Aires, Argentina

*Present address: Cognition and Brain Sciences Unit, Medical Research Council, Cambridge, UK.

Correspondence to: Facundo Manes,
Institute of Cognitive Neurology (INECO),
Castex 3293 (1425), Buenos Aires, Argentina
E-mail: fmanes@neurologiacognitiva.org

Traditional cognitive tests may not be sensitive for the early detection of executive and social cognitive impairments in the behavioural variant of frontotemporal dementia. The aim of this study was to detect specific executive and social cognitive deficits in patients with early behavioural variant frontotemporal dementia using a battery of tests previously shown to be sensitive to frontal lobe dysfunction. Behavioural variant frontotemporal dementia patients and paired controls were assessed with a complete standard neuropsychological battery evaluating attention, memory, visuospatial abilities, language and executive functions. All participants were then assessed with our Executive and Social Cognition Battery, which included Theory of Mind tests (Mind in the Eyes, Faux Pas), the Hotel Task, Multiple Errands Task-hospital version and the Iowa Gambling Task for complex decision-making. Patients were divided into two groups according to their Addenbrooke's Cognitive Examination scores, a measure of general cognitive status. Low Addenbrooke's Cognitive Examination patients differed from controls on most tasks of the standard battery and the Executive and Social Cognition Battery. While high Addenbrooke's Cognitive Examination patients did not differ from controls on most traditional neuropsychological tests, significant differences were found between this high-functioning behavioural variant of frontotemporal dementia group and controls on most measures of our Executive and Social Cognition Battery. Our results suggest that the Executive and Social Cognition Battery used in this study is more sensitive in detecting executive and social cognitive impairment deficits in early behavioural variant of frontotemporal dementia than the classical cognitive measures.

Keywords: frontotemporal dementia; neuropsychological assessment; executive function; frontal lobe

Abbreviations: FTD = frontotemporal dementia; MET = Multiple Errands Test; CDR = Clinical Dementia Severity Rating Scale; ACE = Addenbrooke's Cognitive Examination; MMSE = Mini Mental State Exam; TMT-A = Trail Making Test Part A; FAB = Frontal Assessment Battery; ESCB = Executive and social cognition battery; IGT = Iowa gambling task; ROC = receiver operating characteristic; ACE = Addenbrooke's Cognitive Examination

Received May 11, 2008. Revised January 27, 2009. Accepted January 29, 2009. Advance Access publication March 31, 2009

Introduction

Frontotemporal dementia (FTD) is a degenerative disorder and its clinical manifestations are a direct reflection of the pathological changes that occur in the temporal and frontal cortices (Kipps *et al*., 2007*a*; Rosen *et al*., 2002). Major clinical syndromes may be distinguished: (i) the behavioural variant (bvFTD); (ii) progressive non-fluent aphasia and semantic dementia; and (iii) the motor branch, which includes corticobasal degeneration, progressive supranuclear palsy and motor neuron disease that are associated with FTD features and pathology (Kertesz *et al*., 2005).

Initial symptoms of bvFTD include changes in personality, impaired social interaction, disinhibition, deficits in impulse control and loss of insight (Hodges and Miller, 2001). These patients may also present with compulsiveness, perseverations or stereotyped repetitive acts (Bozeat *et al*., 2000), lack of responsibilities, withdrawal and apathy (Neary *et al*., 1998; Hodges and Miller, 2001). In some patients, an increased appetite with a tendency for sweet foods is present and may be accompanied by hyperorality, though these are usually observed in the later stages of the disease. During the early stages, nonetheless, conventional brain imaging techniques (CT, MRI and SPECT) can be insensitive (Davies *et al*., 2006; Kipps *et al*., 2007*b*; Mendez *et al*., 2007; Rascovsky *et al*., 2007), making diagnosis extremely challenging. This also explains why bvFTD patients are usually misdiagnosed with primary psychiatric syndromes or solely underdiagnosed (Kipps *et al*., 2007; Mendez *et al*., 2007; Rascovsky *et al*., 2007). Moreover, general cognitive functions such as memory, language and praxis may be relatively spared (Gregory and Hodges, 1996; Walker *et al*., 2005). Classical tests of executive function such as the Wisconsin Card Sorting Test, verbal fluency tests or the Trail Making Test may fail to detect the dysexecutive syndrome shown by these early bvFTD patients in everyday life (Gregory *et al*., 1999, 2002). It is now accepted that traditional tasks relying on executives abilities, are not very sensitive to early bvFTD (Hodges, 2007). A recent revision of diagnostic and research criteria for bvFTD was proposed in the hopes of achieving higher shared comparability between research groups (Rascovsky *et al*., 2007) while providing a more comprehensive clinical profile of this disease. Deficits in executive and social cognition tasks are now considered a core feature in its entirety (Kipps and Hodges 2006; Rascovsky *et al*., 2007). In this respect, the presence and detection of executive and social dysfunction may be particularly useful for diagnosis during the early stages of the disease (Hodges, 2007). Although detecting executive deficits does not prove diagnosis of bvFTD, a new battery of executive and social cognition may be valuable in order to consolidate early diagnosis, as established and required by the new criteria (Rascovsky *et al*., 2007). This becomes a crucial point, as a growing number of studies are showing that severe deficits in various aspects of social cognition (i.e. Theory of Mind) and complex decision-making tend to appear in early stages of bvFTD (Gregory *et al*., 2002; Torralva *et al*., 2007). Yet, these domains are not often part of the assessment of patients with early bvFTD.

For this reason, our study attempted to enhance the traditional cognitive assessment battery with the addition of more 'comprehensive' tests of complex executive-social functioning in order to increase sensitivity for the detection of executive and social cognitive impairment in patients with early bvFTD. Despite the fact that some of the tests included in the battery have already been used in patients with traumatic brain injury (Shallice and Burgess, 1991; Manly *et al*., 2002), their usefulness has not been demonstrated in bvFTD.

The new proposed battery consists of tests that measure performance of 'daily life' activities within a 'real life' environment: (i) Multiple Errands Test (Burgess, 2002); (ii) Hotel Task (Manly *et al*., 2002); (iii) complex decision making (Iowa Gambling Task) (Bechara *et al*., 1994); and (iv) social cognition (Theory of Mind tests) (Baron-Cohen, 1997; Stone *et al*., 1998).

(a) Shallice and Burgess (1991) described the MET with the objective of trying to adequately reflect how executive impairments are not reliable predictors of everyday problems manifested in the context of everyday functioning. They demonstrated that patients with frontal lobe damage may be specifically impaired in everyday situations that require planning and multitasking, despite normal performance on tests of language, perception, memory and executive functioning. The authors argued that the MET captures a range of 'real life' activities within the context of a 'real life' environment. More recently, a simplified MET was designed for use in a hospital environment (Burgess, 2002). They found that people with acquired brain injury committed more errors than neurologically healthy controls.

(b) In the area of cognitive rehabilitation, Manly *et al*. (2002) examined whether patients with traumatic brain injury who seemed to experience difficulties in everyday situations would perform more poorly than controls on a modification of the Six Elements Task: 'The Hotel Task'. The central question they addressed was whether the provision of a brief alerting tone during performance would facilitate a closer link between a patient's stated task plan and their actions. The results showed that the group of brain-injured patients performed significantly poorer than age-, gender- and current IQ-matched controls.

(c) Initially, Bechara *et al*. (1994) developed a novel task that simulates real-life decision-making in the way it factors uncertainty of premises and outcomes as well as rewards and punishments. They showed that patients with damage involving orbitofrontal cortex displayed severe impairments in real-life decision making, despite remaining unimpaired intellectually and on traditional neuropsychological measures (Eslinger and Damasio, 1985; Shallice and Burgess, 1991). Convergent evidence confirms the importance of orbitofrontal cortex, but also highlights the relevance of lesion laterality, lesion aetiology and the contribution of other brain regions (including the dorsal prefrontal cortex and amygdala) to decision-making abilities (Manes *et al*., 2002; Clark and Manes, 2004). We have previously reported important deficits in decision making using the Iowa Gambling Task (IGT) test, showing the high sensitivity for the detection of early changes in the prefrontal cortex in bvFTD dementia (Torralva *et al*., 2007).

(d) Early FTD pathology involves the orbitofrontal cortex (Rosen *et al*., 2002), explaining why social cognition tests may be highly sensitive to detect early deficits. Numerous clinical reports have demonstrated the close link between frontal lesions and the genesis of antisocial behaviour, with damage to the orbitofrontal region being particularly implicated (Eslinger, 1999). Stone *et al*. (1998) found that patients with bilateral orbitofrontal lesions performed well on simpler tests of theory of mind, but showed deficits on more subtle social reasoning (faux pas), similarly to individuals with Asperger's syndrome. Functional imaging data support the role of the frontal lobe particularly in theory of mind with activation of the left medial frontal area when considering thoughts and feelings of characters (Fletcher *et al*., 1995) or when reasoning about the beliefs and intentions of others (Goel *et al*., 1995). Moreover, Bora *et al*. (2006) suggest that social cognition is important for real-life functioning. Therefore, the assessment of performance on theory of mind tests in patients with bvFTD is relevant in understanding the changes in their social interaction and, therefore, has implications for social functioning.

Based on the aforesaid statements, the aim of this study was to highlight the usefulness of incorporating more 'comprehensive' tests of executive-social function in neuropsychological assessment battery in order to detect specific cognitive deficits in the earlier phases of bvFTD. In particular, we examined the usefulness of the new assessment battery of executive and social functioning in differentiating groups of high-functioning and low-functioning frontal dementia from non-dementia controls. We hypothesized that the new battery of executive-social functioning tests would more accurately differentiate the high-functioning dementia group from healthy controls relative to 'classical' tests of executive functioning.

Methods

Participants

The procedures were approved and supervised by the ethics committee at our Institute in Buenos Aires, Argentina, and all participants were required to give their informed consent during the initial interview. Thirty-five bvFTD patients with early/mild stages of the disease were recruited, and all patients fulfilled Lund and Manchester criteria for bvFTD diagnosis (Neary *et al*., 1998). Patients presented with prominent changes in personality and social behaviour verified by a caregiver. Patients were assessed at the Raúl Carrea Institute and at INECO in Buenos Aires, Argentina. Dementia severity was assessed using the Clinical Dementia Severity Rating Scale (CDR) (Hughes *et al*., 1982). Diagnosis was initially made by two experts in FTD (F.M. and T.T.). Each patient, individually, was reviewed in the context of a multidisciplinary clinical meeting, where cognitive neurologists, psychiatrists and neuropsychologists discuss each patient's case in particular. BvFTD patients were recruited as part of a broader ongoing study on fronto-temporal dementia. All patients underwent a standard examination battery including neurological, neuropsychiatric and neuropsychological examinations and a MRI-SPECT. They all showed frontal atrophy on MRI, and frontal hypoperfusion on SPECT, when available. Although in the current diagnosis criteria abnormal imaging findings are not necessary, we only included in this study patients with frontal atrophy. Inter-rater reliability for diagnosis was excellent (Cohen's $\kappa = 0.91$). The patients described in the present study did not meet criteria for specific psychiatric disorders.

Their performance was compared to a group of healthy controls ($n = 14$) who were recruited from the same geographical area as the patients and matched for age, gender and levels of education. All participants were assessed with a standard cognitive battery, which included classical executive tests, and the experimental battery designed for the present study. Cognitive assessment was completed in two sessions. Average administration time for this battery was 60 min.

Cognitive assessment

General neuropsychological battery

Cognitive status was measured using the Addenbrooke's Cognitive Examination (ACE) and the Mini Mental State Exam (MMSE; Folstein *et al*., 1975). The ACE is a well-validated scale, which has been shown to be useful for the assessment of patients with dementia (Mathuranath *et al*., 2000). The ACE is also a simple monitoring tool, which can detect progression of disease in FTD (Kipps *et al*., 2008). Premorbid IQ was assessed using the Buenos Aires Word Accentuation Test (WAT-BA; Burin *et al*., 2000). Attention and concentration were assessed with the forward digit span task (Wechsler, 1991) and the Trail Making Test Part A (TMT-A; Partington and Leiter, 1949). Memory was assessed using the logical memory (story recall) subtest from the Wechsler Memory Scale-Revised (Wechsler, 1991). Language comprehension was assessed with the adapted version of the Token Test (Spreen and Benton, 1977) and naming with the adapted version of the Boston Naming Test (Kaplan *et al*., 1983).

Executive function tasks

Executive function tests in the cognitive battery included:

The Frontal Assessment Battery test (FAB; Dubois *et al*., 2000) including six subtests: conceptualization, mental flexibility, motor programming, sensitivity to interference, inhibitory control and environmental autonomy.

Backward digit span: Participants are presented with sequences ranging from two to eight digits in length and are asked to repeat the digits in the reverse order. This task assesses mental manipulation and working memory (Wechsler, 1991).

Letters and Numbers: Participants are presented with an increasing number of letters and digits and are asked to repeat them in a way such that numbers are ordered in an ascending fashion and letters arranged alphabetically. This test also assesses mental manipulation (Wechsler, 1939).

Word fluency tests: The purpose of this test is to assess spontaneous production of words beginning with a given letter or of a given semantic class in a limited amount of time. For letter fluency (phonetic association), participants were asked to verbally produce as many words as possible beginning with a given letter ('P', in this case) for 1 min. For category (semantic association) fluency, participants were asked to produce as many animal nouns as possible within the same time frame.

Trail Making Tests Part B: Participants were asked to join 25 randomly arranged numbers and letters in an alternating fashion.

These tests are designed to assess speed of attention, sequencing, mental flexibility, visual search and set shifting (Partington, 1949).

Modified Wisconsin Card Sorting Test: We used the card sorting version modified by Nelson in which ambiguity is eliminated by removing those cards that share more than one attribute with the stimulus cards. This test assesses abstraction ability and the ability to shift cognitive strategies and is a classic test of executive functions (Nelson, 1976).

Executive and social cognition battery (ESCB): This battery consisted of five tests, including some tests of ecological validity, selected to detect executive-social dysfunction.

Multiple Errands Test Hospital Version (MET-HV): This test, which is frequently administered at the hospital and its surroundings, requires participants to carry out a number of tasks simulating 'real life' situations where minor inconveniences can take place (Burgess, 2002). While still inside the hospital, the patient is given a card with four sets of simple tasks totaling 12 subtasks. The first set requires participants to attain six specific goals, which include purchasing three items (a soda, a postcard and a letter), collecting an envelope from reception, using the internal phone and posting something to an external address. The second set involves obtaining and writing down pieces of information (the area code of the Argentine city 'Chivilcoy', the price of a dinner menu and the last transfer shuttle schedule to Buenos Aires). In the third set, the participant is required to call the proctor 20 min after the test has begun and state the time over the phone. The final task requires the participant to inform the proctor when every task has been completed. Nine rules are clearly stated in the instruction sheet and the participant's behaviour while carrying the tasks is monitored by two observers. At the end of the test, each participant has to indicate on a 10-point scale how well they thought they had done. Errors in this test were categorized as: (i) inefficiencies—where a more effective strategy could have been applied; (ii) rule breaks—where a specific rule (social or one of the nine explicitly defined within the test) was broken; (iii) interpretation failure—where the requirements of a task had been misunderstood; (iv) task failures—where any of the 12 tasks had not been fully completed; and (v) total fails—the sum of all the previous (Shallice and Burgess, 1991).

The Hotel Task: We adapted the task proposed by Manly (Manly *et al*., 2002) for the rehabilitation of executive symptoms, while preserving its most important features. The task comprises six activities that would plausibly need to be completed in the course of running a hotel. The materials needed to perform these activities were arranged on a desk and randomly distributed between participants and sessions. The instructions were as follows: 'In this task you are asked to imagine that you are working in a hotel. Your manager is keen for you to try each of these five everyday activities during the next 15 min so that you can get a 'feel' for the tasks—and make an informed estimate of how long each task would take to complete. Your main goal is to attempt to do each of these five tasks over the next 15 min. There are five main tasks to do. Each of the tasks may take longer than 15 min to complete on its own, so there is no way that you will be able to complete all of them. The most important thing is to try and do a little of each task—spending as much time on each as possible within the total time available'. The details for each of the following tasks were then described. Also, a written summary of the task was placed on top of the relevant materials. The tasks were as follows:

(1) Compiling individual bills. Participants were provided with a group of bills that needed to be arranged by guest name.
(2) Sorting the charity collection. The materials included a box containing 200 coins, which needed to be grouped by country of origin (Argentine, French, Italian, American and Hungarian).
(3) Looking up telephone numbers. Participants were provided with a list of 34 local companies and asked to find and note down their telephone numbers using the regional Yellow Pages phone directory.
(4) Sorting conference labels. Participants were provided with a pile of 100 labels, each with the name of a guest attending a conference. Prior to each administration, the pile was shuffled and participants were asked to sort the cards into alphabetical order.
(5) Proofreading the hotel leaflet. Participants were asked to check a nine-page draft of a proposed new leaflet for the hotel for typogramatical errors.
(6) Opening and closing the garage doors. At two pre-defined times, the participants were asked to remember to open and close the hotel garage doors, in order to allow deliveries. The door was opened by pressing a red button and closed by pressing a black button, both mounted on a single button box placed on the desk.

In addition to the materials needed for each task, a clock was placed on the desk as well. The participant can check the clock at any time, but it is covered in order to assess how frequently the patient needs to watch the clock in order to organize their actions.

The scoring of the Hotel Task was as follows: (i) number of main tasks attempted (out of 5); (ii) number of tasks attempted correctly (out of 5); (iii) time allocation—the optimal allocation was 3 min per task and deviations (in seconds) from this timeframe were calculated and totaled; (iv) number of garage door buttons pressed (out of 2); and (v) garage door time deviations (for detailed scoring, see Manly, 2002).

Iowa gambling task: The computerized version of the IGT mimics real-life personal decision-making activities in real time that include reward and punishment (Bechara *et al*., 1994). Participants are asked to continuously select cards from four decks (A, B, C and D) in order to make as much money as possible in the game. The task is completed after 100 selections. Following card selection, participants receive a certain amount of reward, but some choices also result in loss of money (penalties). Decks A and B are ultimately risky (large rewards and large punishments) while C and D are more conservative (small rewards and small penalties). Under this paradigm, net earnings may only be obtained by consistently selecting from low yield decks (C and D). The dependent variable on this task is the Net Score, calculated by subtracting the number of choices to the risky decks (A + B) from the choices to the safe decks (C + D). In order to quantify the progression of decision-making preference profiles throughout the task, the 100 choices are split into five blocks of 20 consecutive cards. A net score is then calculated for each block. According to previous literature on decision-making cognition showing that the IGT mimics decision-making in real-life scenarios, we purposely included this task to measure real-life decision-making.

The Mind in the Eyes Test: This task consists of 17 photographs of the ocular region of different human faces (Baron-Cohen *et al*., 1997). Participants are required to choose between two options (adjectives) that best describes what the individual is thinking or feeling.

Faux Pas Test: In this test, participants read stories that may contain a social faux pas (Stone *et al*., 1998). After each story, participants are asked whether something inappropriate was said, and if so, asked to give an explanation as to why it was inappropriate. In order to understand that a faux pas has occurred, the participant has to represent two mental states. First, that the person committing the faux pas is unaware that they have said something inappropriate and, second, that the person hearing it might feel hurt or insulted. Each story is presented

in front of the patient in order to decrease working memory load. A memory question is used as a control to check that certain aspects of the stories are retained (Stone *et al*., 1998; Lough *et al*., 2001) and scoring is computed (out of 20 total points) by adding the number of correctly detected faux pas (maximum 10 points) and the number of correctly detected non-faux pas scenarios (maximum 10 points).

Statistical analysis

Based on previous reports defining FTD groups according to cognitive performance (Rahman *et al*., 1999), participants were initially clustered into three groups for statistical analysis: (i) healthy controls (CTR); (ii) bvFTD patients with an ACE score higher than 86 points (hiACE); and (iii) bvFTD patients with an ACE score of 86 or less points (loACE). The score of 86 points (out of a 100) was used, because it is accepted as the cut-off value of the ACE in diagnosing dementia (Mathuranath *et al*., 2000). It is important to note that hiACE patients included in this study were longitudinally studied, and all of them developed clinically and radiologically defined bvFTD. Demographic and neuropsychological data were compared for all three groups using a one-way ANOVA design Tukey's HSD *post hoc* comparisons when appropriate. For variables with a non-homogenous distribution, the Kruskal–Wallis *H* test was used, followed by Mann–Whitney *U* tests to contrast two groups at a time. When analysing categorical variables (e.g. gender and recognition), the Freeman–Halton extension of the Fisher exact probability test for 2×3 contingency tables was used. Inter-rater reliability was determined using Cohen's κ based on the ratings of all 49 participants made by two specialized neuropsychologists who were blind to their diagnosis.

A global score was determined by adding the following individual scores: Mind in the Eyes total score, Faux Pas correct score, total IGT score, number of completed tasks on the Hotel Task and the number of rule breaks on the MET-hv. For the latter, because global score was calculated so that higher values reflected a better performance on the ESCB, and the number of rule breaks on the MET-hv was subtracted rather than added. Global score was used in receiver operating characteristic (ROC) curve analyses to determine the relative sensitivity and specificity of the battery. The area under the ROC curve was determined for the global score and the standard executive tasks. Area under the ROC curve values were compared between each other in order to test statistical differences following Hanley and McNeil (1983) algorithms. The global score and tasks of the ESCB battery were correlated with executive tasks of the standard cognitive battery using Spearman's rank correlation within both FTD groups (hiACE and loACE).

For specific data sets with repeated measures, such as the IGT, the 100 choice were grouped into five blocks of 20 consecutive card, each with a net score for each block calculated as (C + D) − (A + B) decks. An overall IGT score was also determined by adding up the individual block scores. A repeated measures ANOVA 3×5 design was used with group as the between-subjects variable and card blocks as the within-subjects variable. One-way ANOVAs on the card blocks were then conducted, followed by Tukey's HSD *post hoc* tests when significance was reached. The α value for all statistical tests was set at 0.05, two-tailed.

Results

Neuropsychological profile

Clustering of bvFTD patients into groups according to their ACE scores resulted into a group of 16 participants with ACE scores above the cut-off (hiACE) and a group of 19 participants with ACE scores of 86 or below (loACE). General demographic information and neuropsychological test results are summarized for the CTR, hiACE and loACE groups in Table 1.

All groups were successfully matched for age ($F_{2,46} = 1.34$, $P = 0.18$), gender (Fisher's test, $\chi^2_2 = 0.12$, $P = 0.94$) and years of education ($F_{2,46} = 0.53$, $P = 0.95$). As expected, a significant difference was found for ACE scores between the groups ($\chi^2_2 = 35.8$, $P < 0.001$). This difference was significant between the CTR and the loACE group ($U = 2.1$, $P < 0.001$), but not between CTR and the hiACE group ($U = 68.2$, $P = 0.08$). A similar pattern was observed for the MMSE ($\chi^2_2 = 18.5$, $P < 0.001$), with CTR significantly differing from loACE ($U = 26$, $P < 0.001$) but not hiACE ($U = 72.5$, $P = 0.11$).

Attention was significantly different between all groups, as measured both by the forward digit span test ($\chi^2_2 = 17.9$, $P < 0.001$) and the TMT-A ($F_{2,46} = 5.7$, $P < 0.01$). While hiACE patients did not differ significantly from CTR in their forward digits span, TMT-A performance was significantly different between the two groups ($t_{28} = -2.18$, $P = 0.037$). Instead, loACE performance was significantly different from CTR in both tasks ($U = 30.5$, $P < 0.001$ for digits span; $t_{31} = -3.6$, $P = 0.001$ for TMT-A). While group differences were found for all variables of the FAB, none was significantly different between hiACE and CTR, even though loACE differed from CTR (all $P < 0.001$) in similarities ($U = 78.0$), motor series ($U = 58.0$), inhibitory control ($U = 63.0$) and go-no-go ($U = 24.0$).

Significant differences were also found between the three groups in regards to memory performance, including immediate recall ($\chi^2_2 = 19.0$, $P < 0.001$), delayed recall ($\chi^2_2 = 23.4$, $P < 0.001$) and recognition ($\chi^2_2 = 19.5$, $P < 0.001$). Similar to the previously described cognitive domains, hiACE performance did not differ significantly from the CTR group, but loACE significantly differed from all three measures of memory ($U = 29$, $P < 0.001$ for immediate recall; $U = 15.5$, $P < 0.001$ for delayed recall; $U = 29$, $P < 0.001$). Language performance also significantly differed between the groups, as measured by the Token test ($U = 20.0$, $P < 0.001$) and the Boston Naming test ($U = 23.8$, $P < 0.001$). Specifically, performance of hiACE significantly differed from the CTR on the Boston Naming test ($U = 60$, $P = 0.031$) but not on the Token test ($U = 75.5$, $P = 0.13$). The loACE group differed from the CTR, however, on both tests ($U = 10.5$, $P < 0.001$ and $U = 20.5$, $P < 0.001$, respectively).

Performance on every classical executive test included in the battery was significantly different for loACE compared to CTR: digits backward span ($U = 32.5$, $P < 0.001$), Letters and Numbers ($U = 10.5$, $P < 0.001$), phonologic ($U = 43.0$, $P < 0.001$) and semantic ($U = 10.5$, $P < 0.001$) fluency, TMT-B ($t_{31} = -6.8$, $P < 0.001$) and the total score of the WCST ($U = 16.5$, $P < 0.001$) as well as the number of perseverative errors ($U = 30.5$, $P < 0.001$). However, only the Letters and Numbers test ($U = 29.5$, $P < 0.001$) and the TMT-B latency ($t_{28} = -2.24$, $P = 0.033$) were significantly different between hiACE and CTR; instead, digits backward ($U = 80.5$, $P = 0.19$), phonologic ($U = 95.5$, $P = 0.49$) and semantic ($U = 73.5$, $P = 0.11$) fluency and the total score ($U = 70.0$, $P = 0.07$) and number of perseverative errors

Table 1 Demographic information and neuropsychological test performance for the control (CTR), hiACE and loACE groups

	CTR ($n=14$)	hiACE ($n=16$)	loACE ($n=19$)	CTR versus hiACE	CTR versus loACE	hiACE versus loACE
Demographics						
Age (years)	65.5 (6.5)	65.0 (7.4)	69.1 (5.7)	NS	NS	NS
Gender (M:F)	7:7	7:9	9:10	NS	NS	NS
Education (years)	13.9 (3.0)	13.8 (3.8)	13.5 (5.2)	NS	NS	NS
Cognitive status						
ACE	94.5 (5.3)	91.0 (2.6)	74.2 (8.4)	NS	$P<0.001$	$P<0.001$
MMSE	29.2 (1.0)	28.2 (1.9)	25.7 (3.2)	NS	$P<0.001$	$P<0.001$
Attention						
Digit forward span	5.0 (1.1)	4.5 (1.3)	3.3 (1.1)	NS	$P<0.001$	$P<0.001$
TMT-A (s)	39.4 (16.4)	59.4 (30.4)	72.3 (31.4)	$P=0.037$	$P<0.01$	NS
FAB						
Similarities	2.9 (0.3)	2.6 (0.8)	2.3 (0.9)	NS	$P<0.001$	$P<0.01$
Fluency	2.9 (0.3)	2.9 (0.3)	2.0 (1.2)	NS	$P<0.001$	$P<0.001$
Motor series	2.8 (0.4)	2.7 (0.6)	1.7 (1.2)	NS	$P<0.001$	$P<0.001$
Inhibitory control	3.0 (0.0)	2.6 (0.9)	2.0 (1.2)	NS	$P<0.001$	$P<0.01$
Go-no-Go	2.9 (0.3)	2.4 (1.0)	1.1 (1.1)	NS	$P<0.001$	$P<0.001$
Prehension behaviour	3.0 (0.0)	2.9 (0.5)	3.0 (0.0)	NS	NS	NS
Memory						
Immediate	25.1 (8.9)	21.8 (6.7)	12.7 (5.6)	NS	$P<0.001$	$P<0.001$
Delayed	20.1 (8.9)	15.6 (8.1)	4.7 (5.3)	NS	$P<0.001$	$P<0.001$
Recognition	17.1 (2.9)	16.3 (3.2)	10.2 (4.3)	NS	$P<0.001$	$P<0.001$
Language						
Token test	25.2 (1.1)	23.5 (4.4)	20.3 (4.4)	NS	$P<0.001$	$P<0.01$
Boston Naming test	19.8 (0.4)	18.9 (1.2)	16.8 (3.6)	$P=0.031$	$P<0.001$	$P<0.001$
Semantic fluency	20.7 (5.2)	17.8 (5.3)	9.9 (3.8)	NS	$P<0.001$	$P<0.001$
Executive (classic)						
Digit backwards span	5.0 (1.1)	4.5 (1.3)	3.2 (1.1)	NS	$P<0.001$	$P<0.01$
Letters and Numbers	11.1 (2.6)	7.6 (2.1)	5.8 (1.9)	$P<0.001$	$P<0.001$	$P<0.01$
Phonologic fluency	17.5 (5.7)	15.9 (5.5)	9.4 (7.1)	NS	$P<0.001$	$P<0.001$
TMT-B (s)	94.1 (44.3)	145.6 (75)	214 (53)	$P=0.033$	$P<0.001$	$P<0.01$
WCST (total score)	5.6 (0.7)	4.3 (1.6)	2.5 (1.6)	NS	$P<0.001$	$P<0.01$
WCST (pers. errors)	2.2 (2.9)	7.4 (7.3)	11.9 (8.1)	NS	$P<0.001$	$P<0.01$

Values are shown as Mean (SD). ACE = Addenbrooke's Cognitive Examination; MMSE = Mini-Mental State Examination; TMT = Trail Making Test (-A & -B parts); FAB = Frontal Assessment Battery; WCST = Wisconsin Card Sorting Test; NS = not significant.

($U=72.5$, $P=0.25$) of the WCST failed to reliably distinguish the hiACE versus CTR groups.

Executive and social cognition battery

Inter-rater reliability for the ESCB, as measured by the global score, was excellent (Cohen's $\kappa=0.97$). Decision-making performance on the Iowa Gambling Task is depicted in Fig. 1. Data obtained from the IGT task was initially analysed with the 20-block clustering as described in the Methods section. The 3×5 ANOVA analysis revealed no main effect of block ($F_{4,168}=0.882$, $P=0.48$) but a main effect of group ($F_{2,42}=6.44$, $P<0.001$) and a block × group interaction ($F_{8,168}=2.62$, $P=0.03$). One-way ANOVA for repeated measures of block showed no group differences in block 1 ($F_{2,42}=1.66$, $P=0.203$) or block 2 ($F_{2,42}=2.49$, $P=0.095$), while significant differences were found for block 3 ($F_{2,42}=8.81$, $P<0.001$), block 4 ($F_{2,42}=9.81$, $P<0.001$) and block 5 ($F_{2,42}=21.7$, $P<0.001$). *Post hoc* analysis using Tukey's HSD method for these three blocks showed significance resulted from reliable differences between CTR and hiACE ($P<0.05$ for all three blocks) and loACE ($P<0.05$ for all three blocks). The performance of hiACE and loACE was not significantly different at any point during the task.

Performance on tests of theory of mind was significantly different between hiACE and CTR, both in the faux pas test ($U=29.0$, $P<0.001$) and the Mind in the Eyes test ($U=30.0$, $P<0.001$). Three out of the four scoring variables of the MET-HV also revealed significant differences between the hiACE and the CTR groups, including measurements of inefficiencies ($U=26.5$, $P<0.01$), breaking of rules ($U=20.5$, $P<0.001$) and failures to complete the tasks ($U=33.0$, $P<0.01$); failures to interpret the tasks was not reliably different ($U=50.0$, $P=0.12$). In addition, the time deviation measurement of the Hotel Task was significantly different between these two groups ($U=46.0$, $P<0.001$) (Table 2). Performance of loACE patients was significantly different from that of controls on every single variable of the ESCB. Individual patient scores on tasks of the ESCB can be seen in Fig. 2.

One-way ANOVA on the global score showed a significant difference between the groups ($F_{2,42} = 29.667$, $P < 0.001$). Specifically, controls differed from hiACE ($P < 0.001$) and loACE ($P < 0.001$), but the FTD groups did not differ between themselves ($P = 0.521$). A ROC curve between bvFTD and healthy controls showed that a global score of 38 points had a sensitivity of 88.9% and a specificity of 94.3%. The area under the curve was determined for the global score and each of the five tests of the ESCB on ROC analysis. The area under the ROC curve for the global score was 0.975 (CI: 0.934–1.102, $P < 0.001$). Said value was significantly higher than the area under the ROC curve of all five ESCB tests alone (Fig. 3A): versus IGT total score (0.856, $P = 0.047$); versus number of correct tasks on the Hotel task (0.781, $P = 0.034$), versus Faux Pas correct score (0.825, $P = 0.041$), versus MIE (0.752, $P = 0.029$), versus MET-hv rule breaks (0.801, $P = 0.038$). A similar comparison was made between the global score and the measures of executive function of the standard cognitive battery. As shown in Fig. 3B, the global score had a significantly higher area under the ROC curve value than WCST total score (0.827, $P = 0.043$) and preservative errors (0.826, $P = 0.041$), digits backwards span (0.733, $P = 0.021$) and TMT-B (0.160, $P < 0.001$). There was a strong trend to significance between the area under the ROC curve of the global score and Letters and Numbers (0.882, $P = 0.065$).

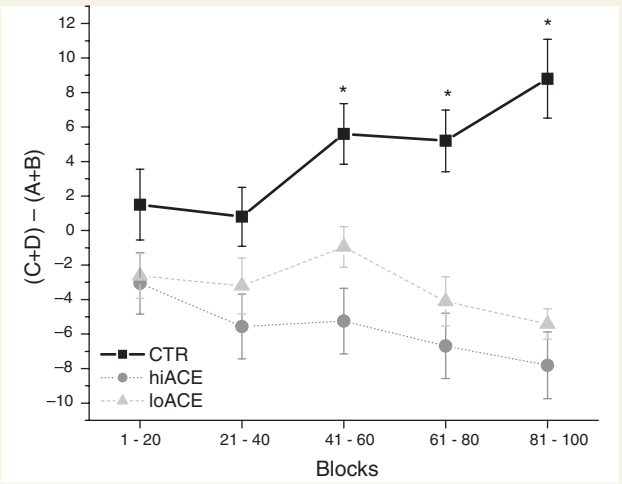

Figure 1 Mean ± SEM scores of the five IGT task blocks for the control, hiACE and loACE groups. Significant differences were observed between the groups on the last three blocks (*$P < 0.001$).

Correlation analyses

The global score did not correlate significantly with any of the executive tasks included in the standard cognitive battery: WCST ($r = -0.273$, $P = 0.113$), WCST perseverative errors ($r = 0.143$, $P = 0.412$), Digit backwards span ($r = -0.153$, $P = 0.379$), Letters and Numbers test ($r = -0.249$, $P = 0.149$), TMT-B ($r = 0.052$, $P = 0.776$) or phonological fluency ($r = 0.178$, $P = 0.306$). A significant correlation was found between the number of correct responses on the Faux Pas and the Mind in the Eyes test ($r = 0.428$, $P = 0.01$). However, neither of these tasks correlated with other variables of the ESCB battery. The Mind in the Eyes task showed a significant correlation with the phonological fluency task ($r = 0.401$, $P = 0.017$), but no significant correlations were found with other executive tasks of the standard battery.

Interestingly, the overall net score of the IGT correlated significantly with the WCST ($r = -0.392$, $P = 0.019$). TMT-B latency significantly correlated with the number of tasks completed on the Hotel Task ($r = -0.351$, $P = 0.037$). No other significant correlations were found between tasks of our ESCB.

In order to examine the relationship of these tasks with memory performance, correlation analyses were conducted between the tasks of the ESCB delayed phase of the logical memory task. No significant correlations were found between the latter and the Mind in the Eyes score ($r = 0.118$, $P = 0.50$), the Faux Pas correct score ($r = 0.127$, $P = 0.47$), the overall IGT net score ($r = -0.248$, $P = 0.15$) and number of rule breaks on the MET-hv ($r = -0.025$, $P = 0.87$). The recall phase of the memory task correlated significantly with the number of tasks completed on the Hotel

Table 2 Executive and Social Cognition Battery (ESCB) performance for the control (CTR), hiACE and loACE groups

	CTR ($n = 14$)	hiACE ($n = 16$)	loACE ($n = 19$)	CTR versus hiACE	CTR versus loACE	hiACE versus loACE
Theory of mind tests						
Faux Pas	19.0 (1.5)	15.4 (2.7)	13.7 (2.2)	$P < 0.001$	$P < 0.001$	NS
Mind in the eyes	14.8 (1.4)	12.6 (1.5)	11.6 (1.9)	$P < 0.001$	$P < 0.001$	NS
MET-HV						
Inefficiencies	0.8 (1.1)	3.4 (2.4)	3.7 (2.4)	$P < 0.01$	$P < 0.001$	NS
Rule breaks	1.0 (1.2)	4.8 (3.3)	4.1 (2.8)	$P < 0.001$	$P < 0.001$	NS
Interpretation failures	0.1 (0.3)	1.6 (2.1)	2.2 (1.8)	NS	$P < 0.001$	NS
Task failures	0.2 (0.4)	2.4 (2.8)	4.4 (3.3)	$P < 0.01$	$P < 0.001$	NS
Hotel Task						
Tasks attempted	4.5 (0.5)	3.9 (1.1)	2.7 (1.0)	NS	$P < 0.001$	$P < 0.01$
Tasks correct	4.4 (0.5)	3.5 (1.5)	2.4 (1.1)	NS	$P < 0.001$	$P = 0.02$
Time deviation	277 (130)	525 (264)	695 (281)	$P < 0.001$	$P < 0.001$	NS
Button pressing	1.9 (0.3)	1.9 (1.8)	0.2 (0.5)	NS	$P < 0.001$	$P < 0.01$
Garage time deviation	4.1 (2.4)	3.6 (2.9)	0.5 (1.0)	NS	$P < 0.001$	$P < 0.01$

Values are shown as Mean (SD). MET-HV = Multiple Errands Task Hospital Version; NS = not significant.

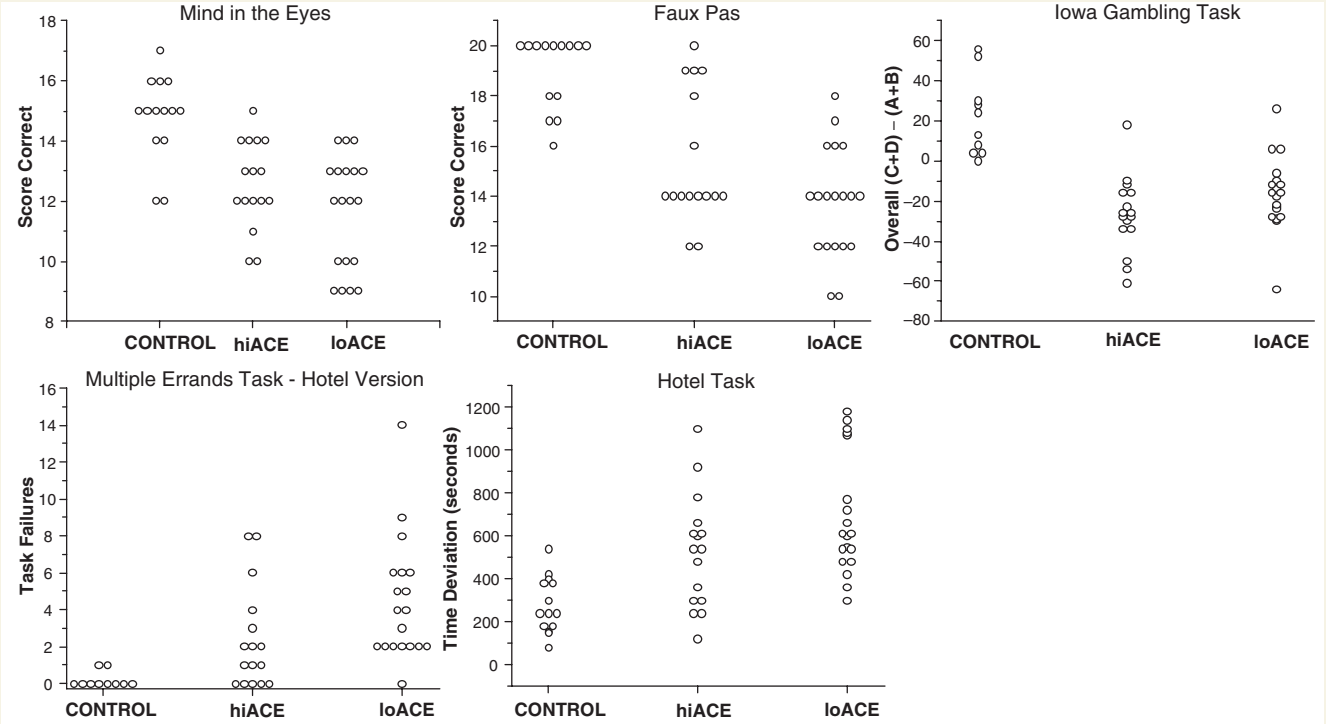

Figure 2 Individual patient scores on the tasks of the Executive and Social Cognition Battery (ESCB).

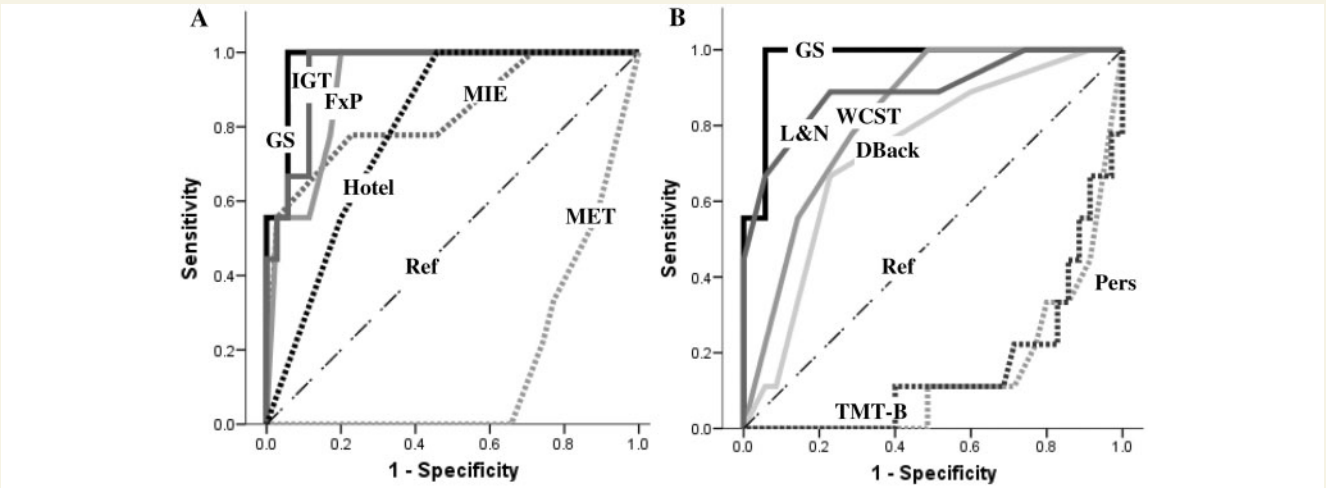

Figure 3 ROC curve analysis for the Global score (GS) and the tasks of (**A**) the ESCB (IGT = Iowa Gambling Task total score; FxP = Faux Pas correct score; MIE = Mind in the Eyes correct score; MET = Multiple Errands Task rule breaks score; and (**B**) the executive tasks of the standard cognitive battery (L&N = Letters and Numbers total score; WCST = Wisconsin Card Sorting Test; DBack = Digits backward span; TMT-B = Trail Making Test Part B latency; Pers = WCST perseverative errors). The reference line is provided (Ref). In both cases, the global score had a larger area under the curve value than the area under the curve of each of the tasks alone.

task ($r = 0.378$, $P = 0.025$). No significant correlations were found with the immediate and recognition phases of memory.

Discussion

We assessed 35 patients with established diagnosis of behavioural variant FTD using a battery (ESCB) that included tests of real-life executive functioning, complex decision-making and social cognition. We determined whether these tests were more sensitive than classical tests of executive function to detect executive and social cognition deficits in the earlier stages of bvFTD. By classifying bvFTD as being either above (hiACE) or below (loACE) the cut-off score of the Addenbrooke's Cognitive Examination (ACE), we were able to show that the hiACE group performed

significantly below healthy controls on the tests of the battery we propose, but did not differ on most of the standard tests.

An emerging problem with many classical tests of executive functions is their lack of real-life or ecological validity. Many frontal patients exhibit planning impairment in real-life situations, despite demonstrating adequate performance on traditional assessment measures. For example, the famous frontal patient E.V.R., described by Eslinger and Damasio (1985) displayed normal to superior performance on standard tests of executive functioning and yet was severely impaired in daily tasks, particularly those that required the executive skills of planning, decision-making and judgment. As emphasized by Gioia and Isquith (2004), on classical neuropsychological tests, the examiner provides the structure, organization, guidance, planning and monitoring necessary for optimal performance, therefore transforming him or herself in the patient's own executive system. Thus, traditional testing environments may fail to induce executive deficits, making assessment of this cognitive domain particularly challenging. Consequently, these tests may not be sufficiently sensitive (Gregory and Hodges, 1996; Gregory *et al*., 1999, 2002) to detect deficits in bvFTD, particularly in earlier stages of the disease. If the patient's environment poses little demand on certain skills, executive deficits may have no impact on real-life settings. On the contrary, otherwise minor executive deficits can become especially impairing in highly demanding environments. The development of assessment tools that mimic real-life scenarios must focus on the detection of the real cognitive demand involved in everyday real-life settings.

In clinical practice, being able to quantify the extent of executive dysfunction in patients who present standard neuropsychological evaluation within normal values while caregivers report severe changes in their behaviour and real-life executive deficits will help with early diagnosis. More sensitive and specific neuropsychological tests like the ones used in this study have the potential to be the most realistic and cost-efficient way of contributing to early diagnosis, by providing a further tool to detect subtle yet relevant changes to cognitive functioning. It would also greatly facilitate the design of appropriate rehabilitation strategies with the objective of improving the impact of these deficits in patients' daily living.

Initial clustering of bvFTD patients into a high-functioning (hiACE) and a low-functioning (loACE) group according to the dementia cut-off score of the Addenbrooke's Cognitive Examination was done based on previous studies of FTD (Rahman *et al*., 1999) that divided patients according to cognitive performance. While Rahman *et al*. (1999) defined mild FTD for patients with a MMSE score of >20 our hiACE patients had MMSE >26. We prefer the ACE over the MMSE as a way to readily classify patients based on a general cognitive status screening tool that has been previously validated (Mathmuranath *et al*., 2000) in this patient population and is extensively used in the field. The ACE has been shown to be effective in measuring progression of disease in FTD without the need of longitudinal neuroimaging (Kipps *et al*., 2008). Under this group division, healthy controls were compared with both dementia groups, but only the loACE differed significantly from the control group on a traditional neuropsychological battery. The fact that the hiACE group showed, for the most part, a performance within normal range, highlights the low sensitivity of traditional cognitive batteries for the detection of subtle cognitive impairments in this patient population. Our findings are consistent with the overall cognitive functioning within normal ranges demonstrated by previous research in early FTD (Hodges, 2007).

Therefore, in trying to design a battery that would increase sensitivity for the detection of these subtle deficits that characterize the bvFTD, we decided to incorporate tasks that would resemble real-life demands, decision-making processes or social situations. As expected, the loACE group differed from healthy controls on all measures of the ESCB. Similarly, the hiACE group differed from controls in measures of real-life planning and organization, complex decision-making and theory of mind tests. On the MET-hv, this group made more errors than healthy controls (inefficiencies), acting more impulsively (rule breaks), with no apparent planning, and poor organization of the tasks (task failures). They differed significantly from healthy controls in the optimal time deviation of the Hotel task, which is a measure closely associated with planning and flexibility, two of the hallmarks of executive functioning. As well, this patient group showed severe deficits in decision-making early (blocks 3, 4 and 5) on the IGT and differed from controls on both tasks of theory of mind.

In accordance, the global score was thought as a measure of performance across all five tasks of the ESCB and was used throughout the analyses in order to interpret psychometric properties of the battery and explore its potential utility. Indeed, the global score differed between controls and both dementia groups, but the latter did not differ between themselves. This trend represented the previously described overall profile we observed on the ESCB. Given the diverse nature of the ESCB tasks, sensitivity and specificity of the battery were assessed based on the global score, obtaining both high sensitivity and specificity for the detection of executive and social cognition deficits. Area under the ROC curve on the global score was interpreted as a measure of accuracy in the detection of cognitive deficits. Thus, the fact that the area under the ROC curve for the global score was higher than that for all five tests independently suggests that administering all tests of the ESCB is more accurate in the detection of frontal symptoms than administering each of the tests alone. Future studies should look at possible combinations of tests that would allow for an abbreviated version of the ESCB. Similarly, administering ESCB is more accurate than assessing executive functions with traditional tests.

There are some limitations to this study. First, diagnosis of bvFTD patients was based on clinical assessment alone and not on pathological analyses. Second, this study only tested specificity of the global score in discriminating bvFTD from normal controls, but future studies should also assess its discriminative power versus other forms of dementia and psychiatric disorders in a sample representative of the total population of patients with behavioural symptoms. As outlined earlier, the objective of this battery was to detect subtle executive and social cognition deficits, therefore contributing to the early diagnosis of bvFTD. Although still early to determine, this battery seems to be sensitive to the cognitive and social components of the early stages of bvFTD.

Clinically, it is important to find a battery of tests capable of detecting early changes in the cognitive profile of bvFTD patients.

As demonstrated in this study, traditional measures of frontal executive function may be ineffective in the early detection of subtle executive and social cognition deficits in bvFTD. Patients with an apparent high performance on a standard neuropsychological test do not differ significantly from controls in their performance on basic cognitive domains or classical tests of executive function. Although detecting executive deficits does not prove diagnosis of bvFTD, this battery may be valuable in order to consolidate early diagnosis as required by the new criteria (Rascovsky *et al*., 2007). A bvFTD patient may go undiagnosed until well into the dementing process, making early diagnosis essential to help families who are faced with relatives behaving bizarrely, make financial errors, developing obsessive behaviours, acting out or behaving sexually inappropriate toward others, with no insight and normal performance on mental status screening tests. Finally, recording early executive dysfunctions can also have legal and ethical implications, as persons with established bvFTD may score well on standard cognitive tasks that are used to determine competence in the legal field. For the earlier stages, while patients still perform normally on most neuropsychological tests, we suggest that this new battery will increase sensitivity in the detection of subtle, yet important, executive and social cognition problems.

Acknowledgements

We thank Agustina Lacroze for testing patients. We also thank Chris Kipps for his helpful comments.

Funding

INECO grant and a Fundación LyD grant.

References

Baron-Cohen S, Jolliffe T, Mortimire C, Robertson M. Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger Syndrome. J Child Psychol Psychiatry 1997; 38: 813–22.

Bechara A, Damasio A, Damasio H, Anderson S. Insensivity to future consequences following damage to human prefrontal cortex. Cognition 1994; 50: 7–15.

Bora E, Eryavuz A, Kayahan B, Sungu G, Veznedaroglu B. Social functioning, theory of mind and neurocognition in outpatients with schizophrenia; mental state decoding may be a better predictor of social functioning than mental state reasoning. Psychiatry Res 2006; 145: 95–103.

Bozeat S, Gregory C, Lambon Ralph M, Hodges JR. Which neuropsychiatric features distinguish frontal and temporal variants of frontotemporal dementia from Alzheimer's disease? J Neurol Neurosurg Psychiatry 2000; 69: 178–86.

Burgess P. Development of a simplified version of the multiple errands test for use in hospital settings. Neuropsychol Rehabil 2002; 12: 231–55.

Burin DI, Jorge RE, Arizaga RA, Paulsen JS. Estimation of premorbid intelligence: the word accentuation test—Buenos Aires version. J Clin Exp Neuropsychol 2000; 22: 677–85.

Clark L, Manes F. Social and emotional decision-making following frontal lobe injury (Review). Neurocase 2004; 10: 398–403.

Davies RR, Kipps CM, Mitchell J, Kril JJ, Halliday GM, Hodges JR. Progression in frontotemporal dementia: identifying a benign behavioral variant by magnetic resonance imaging. Arch Neurol 2006; 63: 1627–31.

Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology 2000; 55: 1621–6.

Eslinger PJ. Orbital frontal cortex: historical and contemporary views about its behavioural and physiological significance. An introduction to special topic papers, Part 1. Neurocase 1999; 5: 225–9.

Eslinger PJ, Damasio AR. Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. Neurology 1985; 35: 1731–41.

Fletcher PC, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RS et al. Other minds in the brain: a functional imaging study of ''theory of mind'' in story comprehension. Cognition 1995; 57: 109–28.

Folstein MF, Folstein SE, McHugh PR. ''Mini Mental State'': a practical method for grading the mental state of patients for the clinitian. J Psychiatr Res 1975; 12: 189–98.

Gioia GA, Isquith PK. Ecological assessment of executive function in traumatic brain injury. Dev Neuropsychol 2004; 25: 135–58.

Goel V, Grafman J, Sadato N, Hallett M. Modeling other minds. Neuroreport 1995; 6: 1741–6.

Gregory CA, Hodges JR. Frontotemporal dementia: use of consensus criteria and prevalence of psychiatric features. Neuropsychiatry Neuropsychol Behav Neurol 1996; 9: 145–53.

Gregory CA, Lough S, Stone V, Erzinclioglu SH, Martin L, Baron-Cohen S, et al. Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. Brain 2002; 125: 752–64.

Gregory CA, Serra-Mestres J, Hodges JR. Early diagnosis of the frontal variant of frontotemporal dementia: how sensitive are standard neuroimaging and neuropsychologic tests? Neuropsychiatry Neuropsychol Behav Neurol 1999; 12: 128–35.

Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839–43.

Hodges JR. Chapter 1: overview of frontotemporal dementia. In: John R Hodges, editor. The Frontotemporal Dementia Syndromes. Cambridge, England: Cambridge University Press; 2007.

Hodges JR, Miller B. Classification, genetics and neuropathology of frontotemporal dementia. Introduction to the special topics papers, part I [Review]. Neurocase 2001; 7: 31–5.

Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982; 140: 566–72.

Kaplan E, Goodglass H, Weintraub S. Boston naming. Test. Philadelphia: Lea & Febiger; 1983.

Kertesz A, McMonagle P, Blair M, Davidson W, Munoz DG. The evolution and pathology of frontotemporal dementia. Brain 2005; 128: 1996–2005.

Kipps CM, Davies RR, Mitchell J, Kril JJ, Halliday GM, Hodges JR. Clinical significance of lobar atrophy in frontotemporal dementia: application of an MRI visual rating scale. Dement Geriatr Cogn Disord 2007b; 23: 334–42.

Kipps CM, Hodges JR. Theory of mind in frontotemporal dementia. Soc Neurosci 2006; 1: 235–44.

Kipps CM, Nestor PJ, Dawson CE, Mitchell J, Hodges JR. Measuring progression in frontotemporal dementia: implications for therapeutic interventions. Neurology 2008; 70: 2046–52.

Kipps CM, Nestor PJ, Fryer TD, Hodges JR. Behavioural variant frontotemporal dementia: not all it seems? Neurocase 2007a; 13: 237–47.

Lough S, Gegory C, Hodges J. Dissociation of social cognition and executive function in frontal variant frontotemporal dementia. Neurocase 2001; 7: 123–30.

Manes F, Sahakian B, Rogers R, Antoun N, Aitken M, Robbins T. Decision-making processes following damage to the prefrontal cortex. Brain 2002; 125: 624–39.

Manly T, Hawkins K, Evans J, Woldt K, Robertson IH. Rehabilitation of executive function: a facilitation of effective goal management on complex tasks using periodic auditory alerts. Neuropsychologia 2002; 40: 2671–81.

Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and fronto-temporal dementia. Neurology 2000; 55: 1613–20.

Mendez MF, Shapira JS, McMurtray A, Licht E, Miller BL. Accuracy of the clinical evaluation for frontotemporal dementia. Arch Neurol 2007; 64: 830–5.

Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 1998; 51: 1546–54.

Nelson H. A modified card sorting response sensitive to frontal lobe defects. Cortex 1976; 12: 313–24.

Partington JE, Leiter RG. Partington's pathway test. Psychol Serv Bull 1949; 1: 9–20.

Rahman S, Sahakian BJ, Hodges JR, Rogers RD, Robbins TW. Specific cognitive deficits in mild frontal variant frontotemporal dementia. Brain 1999; 122: 1469–93.

Rascovsky K, Hodges JR, Kipps CM, Johnson JK, Seeley WW, Mendez MF, et al. Diagnostic criteria for the behavioral variant of frontotemporal dementia (bvFTD): current limitations and future directions [Review]. Alzheimer Dis Assoc Disord 2007; 21: S14–8.

Rosen HJ, Gorno-Tempini ML, Goldman WP, Perry RJ, Schuff N, Weiner M, et al. Patterns of brain atrophy in frontotemporal dementia and semantic dementia. Neurology 2002; 58: 198–208.

Shallice T, Burgess PW. Deficits in strategy application following frontal lobe damage in man. Brain 1991; 114: 727–41.

Spreen FO, Benton AL. Manual of instructions for the Neurosensory Center Comprehensive Examination for Aphasia. Victoria: British Columbia; 1977.

Stone VE, Cohen SB, Knight RT. Frontal lobe contribution to theory of mind. J Cogn Neurosci 1998; 10: 640–56.

Torralva T, Kipps CM, Hodges JR, Clark L, Bekinschtein T, Roca M, et al. The relationship between affective decision-making and theory of mind in the frontal variant of fronto-temporal dementia. Neuropsychologia 2007; 45: 342–9.

Walker AJ, Meares S, Sachdev PS, Brodaty H. The differentiation of mild frontotemporal dementia from Alzheimer's disease and healthy aging by neuropsychological tests. Int Psychogeriatr 2005; 17: 57–68.

Wechsler D. The measurement of adult intelligence. Baltimore: Williams & Wilkins; 1939.

Wechsler D. Wechsler Intelligence Scales for Children. 3rd edn. San Antonio, TX: The Psychological Corporation; 1991.