Journal of the International Neuropsychological Society (2010), Page 1 of 8.

doi:10.1017/S1355617710000482

Utility of an abbreviated version of the executive and social cognition battery in the detection of executive deficits in early behavioral variant frontotemporal dementia patients

EZEQUIEL GLEICHGERRCHT,[1] TERESA TORRALVA,[1,2] MARÍA ROCA,[1,2] AND FACUNDO MANES[1,2]

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

(Received November 10, 2009; Final Revision March 25, 2010; Accepted March 26, 2010)

Abstract

The detection of executive deficits in early behavioral variant frontotemporal dementia (bvFTD) is crucial, as impairments in the executive domain constitute an important diagnostic feature of the newly proposed diagnostic criteria for bvFTD. Our group has recently demonstrated that classical executive tests fail to detect the executive deficits of a subgroup of early bvFTD patients. When administered an executive and social cognition battery (ESCB) that includes tasks that mimic everyday scenarios (e.g., affective decision-making, planning and organization, theory of mind), however, the performance of those bvFTD patients differed significantly from that of controls. One limitation of the ESCB is its lengthy nature (approximately 90 min). For this reason, the present study analyzes the usefulness of alternative shorter versions of this battery. We propose one particular two-task combination that demands approximately 30 min for its administration and scoring, and which presents similar discriminatory accuracy as that of the complete ESCB, while maintaining its significantly superior capacity to detect subtle executive deficits in bvFTD patients relative to classical executive tests. We suggest that, in clinical settings where tools, time, or human resources are scarce, this abbreviated ESCB may be useful in the detection of subtle yet impairing executive impairments of patients with bvFTD. (*JINS*, 2010, 1–8.)

Keywords: Frontal lobe, Prefrontal cortex, Neuropsychological assessment, Ecological assessment, Iowa Gambling Task, Hotel Task

INTRODUCTION

From a clinician's perspective, it can very frustrating to evaluate a patient with a convincing history of executive and social cognition deficits (regardless of whether these have emerged because of early dementia, head injury, encephalitis, or stroke), particularly as described by family members, and yet, observe minimally impaired or within normal performance on formal neuropsychological testing using the standard measures. The family shares this frustration, as they are keenly aware that something is wrong, yet no "objective" evidence of impairment can be found on comprehensive assessments. Such patients are typically labeled as having primary psychiatric disorder, or malingering, or the family would be suspected of distorting the facts for some other gain. The challenge with any neuropsychological battery is finding the best balance between sensitivity/specificity and duration/intensity of testing. The ideal battery should achieve excellent sensitivity and should be performed in a reasonable period of testing time.

Frontotemporal dementia (FTD) is currently conceived as an umbrella term that encompasses several neurodegenerative syndromes (for review, *see* Josephs, 2008). One such disorder is the behavioral variant FTD (bvFTD), which results from prominent changes within the frontal cortex (Forman et al., 2006). Clinically, the bvFTD is evidenced through the behavioral disturbances that characterize the syndrome, which include—even during the earliest stages—altered social interaction, typically exhibiting disinhibition, deficits in impulse control, loss of insight, lack of responsibilities, or even withdrawal and apathy (Neary et al., 1998; Hodges & Miller, 2001). Behavioral disturbances may also present as compulsive behavior, perseverations or stereotyped and repetitive acts (Bozeat, Gregory, Ralph, & Hodges, 2000). Moreover, the neuropsychological profile of bvFTD patients

Correspondence and reprint requests to: Facundo Manes, Castex 3293, Buenos Aires, Argentina. E-mail: fmanes@ineco.org.ar

includes executive/generation deficits with relative sparing of memory and visuospatial functions (Hodges & Miller, 2001; Kipps, Knibb, Patterson, & Hodges, 2008; Neary et al., 1998).

Noticeably, behavioral and neuropsychological changes may occur well before the appearance of any abnormalities on structural neuroimaging (Davies, Kipps, Mitchell, Kril, Halliday, & Hodges, 2006; Kipps, Nestor, Fryer, & Hodges, 2007; Mendez, Shapira, McMurtray, Licht, & Miller, 2007; Rascovsky et al., 2007), which can delay early diagnosis due to under- or misdiagnosis because of overlapping symptomatic profiles with psychiatric disorders. In this sense, developing new tools with increased sensitivity for the detection of initial executive deficits in bvFTD patients is essential, as executive impairment is one of the key features of the new Diagnostic and Research Criteria for bvFTD (Rascovsky et al., 2007). Our group has recently contributed to this goal (Torralva, Roca, Gleichgerrcht, Bekinschtein, & Manes, 2009) by developing an executive and social cognition battery (ESCB) that aims at increasing sensitivity for the detection of executive and social cognition deficits by testing patients with tasks that mimic real-life scenarios more closely than standard tests of executive functioning.

The ESCB is comprised of five tests that measure (a) performance on daily life activities using the Multiple Errands Test (Shallice & Burgess, 1991) and the Hotel Task (Manly, Hawkins, Evans, Woldt, & Robertson, 2002); (b) affective decision-making using the Iowa Gambling Task (Bechara, Damasio, Damasio, & Anderson, 1994); and (c) social cognition using the Reading the Mind in the Eyes test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) and the Faux Pas task (Stone, Baron-Cohen, Knight, 1998). In our original study (Torralva et al., 2009), we assessed a group of early-mild bvFTD patients and controls using screening tests of general cognitive status, a comprehensive standard neuropsychological battery, and the ESCB. Based on whether bvFTD patients scored above or below the cutoff score of one particular screening tests of general cognitive status, the widely-used Addenbrooke's Cognitive Examination (Mathuranath, Nestor, Berrios, Rakowicz, & Hodges, 2000), we divided bvFTD patients into a high-functioning (hiFTD) and a low-functioning (loFTD) group, respectively. As expected, the loFTD group differed significantly from controls on most of the tasks included in the standard battery. Of interest, though, the hiFTD group showed no significant differences from controls on most classical tasks of the battery, and in particular, little differences were found between hiFTD patients and controls on standard tests of executive functioning. Yet, when the scores on the ESCB were compared between the groups, both loFTD and hiFTD patients showed significant differences from controls on all variables of the novel battery we proposed. This finding was particularly important, because it revealed that there is a subset of early bvFTD patients whose early changes may go undetected when assessed with comprehensive standard neuropsychological batteries. Still, tasks that recreate more real-life scenarios were able to detect the subtle yet impairing deficits that characterize this patient population. Comparison of the ESCB's discriminatory accuracy between bvFTD patients and controls was significantly superior to the capacity of classical executive tests to differentiate between the groups.

Besides the great utility of the ESCB from both a clinical and research perspective, administering all five tests can be cumbersome in fast-paced clinical settings, as the comprehensive administration of the ESCB demands over 80 min for instructions, task administration, and scoring. Furthermore, the need for trained neuropsychologists and several stimuli for administration of the tasks make it challenging to administer the complete ESCB in all kinds of settings. As well, there is an increasing need for short yet effective tools to detect subtle cognitive deficits in neurological and neuropsychiatric populations, which may contribute to early diagnosis while keeping costs to a minimum. This, in turn, allows for cognitive assessment to be made available to more patients, which is essential as we gain more evidence of the usefulness of neuropsychology in differential diagnosis and in the design of nonpharmacological treatment plans. For this reason, the present study seeks to investigate the utility of abbreviated versions of the ESCB by comparing their discriminatory accuracy between bvFTD patients and controls.

METHOD

Participants

Patients with diagnosis of bvFTD (n = 35) and controls (n = 14) for this study were part of the sample used in the original publication of the ECSB (Torralva et al., 2009). Diagnosis was initially made by two experts in FTD (F.M. and T.T.). Each patient was individually 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 frontotemporal dementia. All presented with prominent changes in personality plus social behavior verified by a caregiver. FTD diagnosis was made on the basis of published criteria (Neary et al., 1998). 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 criteria diagnosis abnormal imaging findings is not mandatory, we included for this study only patients with frontal atrophy. The patients described in the present study did not meet criteria for specific psychiatric disorders. Patients were in the mild stages of the disease, as determined by a score of 0.5 or 1 on the Clinical Dementia Severity Rating Scale (CDR) (Hughes, Berg, Danziger, Coben, & Martin, 1982), Inter-reliability diagnosis between two experts (F.M. and T.T.) was excellent (Cohen's kappa = .91). Healthy controls were matched for age, gender, and years of education, and they reported no history of traumatic brain injury, psychiatric disorders, or substance abuse. All participants gave their informed consent before inclusion in this study. Further details of the patient population and the control group can be found elsewhere (Torralva et al., 2009).

Procedure

The study was initially approved by the ethics committee at the Institute of Cognitive Neurology (Buenos Aires, Argentina) following the ethical standards established by the 1964 Declaration of Helsinki. BvFTD patients and healthy controls completed a series of interviews, including neurological and psychiatric assessment, standard neuropsychological assessment, and they were all administered the ECSB. For the purposes of the present study, data was obtained from (i) measures of general cognitive status screening, which included the Mini Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) and Addenbrooke's Cognitive Examination (ACE) (Mathuranath et al., 2000); (ii) classical measures of executive functioning, which included backward digit span (BackDS) (Wechsler, 1997), phonological fluency (letter "P", in this case) (PhFlu) (Lezak, Howieson, & Loring, 2004), Trail Making Test Part B (TMT-B) (Partington & Leiter, 1949), and the Wisconsin Card Sorting Test (WCST) (Nelson, 1976); and (iii) measures of the ECSB, which included the Iowa Gambling Task (IGT) (Bechara et al., 1994), the Hotel task (HOT) (Manly et al., 2002), the Multiple Errands Task (MET) (Shallice & Burgess, 1991), the Mind in the Eyes task (MIE) (Baron-Cohen et al., 2001), and the Faux Pas test (FAUX) (Stone et al., 1998). Average time to administer each task (including instructions and scoring) was calculated for all bvFTD patients. The detailed description of these tasks, the rationale for their use, scoring, and interpretation of their results is thoroughly described elsewhere (Torralva et al., 2009).

Statistical Analysis

As shown in Table 1, we chose one representative sub-variable from each of the five tasks in the ESCB, based on our group's previous report of a composite global score calculated by adding up said variables (Torralva et al., 2009). These variables were originally chosen because they were the most frequently reported variables for each test, and they were easy to calculate during patient assessment, therefore minimizing task administration time. For the purposes of the present study, however, to increase accuracy of composite scores, we identified the range of scores $[a_i; b_i]$ that these variables could adopt based on the structure of the task (e.g., number of stimuli), and we determined individual transformed scores (TS) for the five variables based on the following procedure:

Let $X_1 \ldots X_n$ be the variables to be combined for analysis with range $x_i \in [a_i; b_i] \; \forall \; 1 \leq i \leq n$, then

$$TS_i = \frac{|a_i| + j_i}{|a_i| + |b_i|} \text{ where } \begin{cases} j_i = x_i \text{ if } x_i \text{ adopts higher values with better performance} \\ j_i = b_i - x_i \text{ if } x_i \text{ adopts higher values with worse performance} \end{cases}$$

Individual TSs were calculated alone $\binom{1}{5}$ and combined in numbers of two $\binom{2}{5}$, three $\binom{3}{5}$, four $\binom{4}{5}$, and five $\binom{5}{5}$, to determine a "combined mean" ($\hat{C}$) for each of the 31 $\left[\binom{1}{5}+\binom{1}{5}+\binom{1}{5}+\binom{1}{5}+\binom{1}{5}=1+5+10+10+5\right]$ possible combinations based on the formula $\hat{C} = \dfrac{\Sigma \, TS\, combined}{number\ of\ variables\ included}$ such that $0 \leq \hat{C} \leq 1$. In sum, the transformed score for each task is the relation between the individual score on that task (j_i) plus the minimum possible score (a_i) and the sum of the minimum (a_i) and maximum (b_i) possible scores that one may get on the task. This procedure was preferred over other methods to transform scores for several reasons. First, because the five tasks included as part of the ESCB have different ranges of scores (e.g., IGT can take values from −20 to 20 but FAUX can only take values from 0 to 20), the formula described above ensures that the score of one task can be comparable with performance scores from another task of the battery. This is so because when calculating the individual score for a given task, it will adopt values between 0 (worst performance) and 1 (best performance) in all cases, despite the tasks minimum and maximum values and whether higher scores reflect better or poorer performance. This could also be potentially achieved with *z* scores, but unfortunately, no normative data at the large scale has been generated for these tasks. In this sense, computing *z* scores for patients would have required using performance scores of

Table 1. Sub-variables of the tasks administered in the ESCB chosen for combinatory analysis

ESCB task	Sub-variable	Range of values: *Min* (a_i)	*Max* (b_i)	Administration time *Mean* ± *SD*
HOT	Number of tasks completed	0	5	18.7 ± 2.1
MET	Number of rule breaks	0	14	31.0 ± 4.1
IGT	Net score on block 5	−20	20	13.4 ± 3.1
MIE	Total score	0	17	5.1 ± 1.3
FAUX	Total score	0	20	21.0 ± 2.4

Note. Minimum and maximum possible values were determined based on the structure of the task (e.g., number of stimuli). Average administration time (for bvFTD patients, including scoring) is shown.

control participants, and no transformed scores for the healthy population of our study would have been available. For this reason, another advantage of using the aforementioned score transformation procedure is that it generates data for both groups in this study, which is essential in assessing discriminatory accuracies of different task combinations. Moreover, using this procedure also allows for averaging performance across different tasks, and more importantly, across different numbers of tasks (e.g., a two-task *vs*. a five-task combination), as the combined mean will not be biased toward one particular task because all scores will weigh the same to the formula.

Then, $\hat{C}$ values for all combinations were analyzed with ROC curves to calculate the area under the curve (AuC) as a measure of discriminatory accuracy between bvFTD patients and controls. The AuC values for different $\hat{C}$s were compared between each other using the Hanley & McNeil (Hanley & McNeil, 1983) method for ROC curves derived from the same cases. The combination identified as bearing the highest sensitivity/specificity associated with the shortest administration time was used as the independent variable in discriminatory analysis between bvFTD patients and controls. Associated with this analysis, leave-one-out cross-validation (LOOCV) was conducted to assess the generalizability of the results in future patient populations. In LOOCV, one single case is used as the validation sample, while the remaining k-1 cases are used as the training data, and the procedure is repeated until all cases have been validated.

RESULTS

Patients and controls were matched for demographic variables, and no significant differences were found for their age ($t_{47} = -1.37$; $p = .18$), gender ($\chi^2 = 0.01$; $p = .93$), or years of education ($t_{47} = 0.23$; $p = .82$). Performance on the general cognitive status screening tests, on the classical executive function tasks, as well as on the ESCB are presented in Table 2. Comparison analyses across the groups have been described in detail elsewhere (Torralva et al., 2009).

Table 3 reveals that the complete battery (IGT+HOT+MET+MIE+FAUX) had an AuC of 0.981 (*SE* = 0.017). Naturally, this combination had the highest mean administration time (89.2 min). In fact, when the FAUX was not administered (IGT+HOT+MET+MIE), time was reduced by 21 min and discriminatory accuracy increased by 0.9%, making this the most sensitive combination for the detection of bvFTD patients. The two-task combination associated with the highest AuC (0.963; *SE* = 0.028) was IGT+HOT. The discriminatory accuracy of this combination was compared with that of the top five AuC-ranked combinations (Table 4a). No significant differences were found between these top combinations in their capacity to discriminate bvFTD patients from controls. Noticeably, the difference was not significant between this two-task combination and the entire ESCB ($z = 0.51$; $p = .61$). Nor was the difference significant between IGT+HOT and the top-ranked IGT+HOT+MET+MIE combination ($z = 0.88$; $p = .37$). On the other hand, when the discriminatory accuracy of IGT+HOT was compared with that of the bottom five AuC-ranked combinations (Table 4b), significant differences were indeed found with MET ($z = 1.92$; $p = .039$), HOT ($z = 1.87$; $p = .042$), MIE ($z = 1.75$; $p = .049$), and HOT+MIE ($z = 1.71$; $p = .050$), and a strong trend to significance was observed when compared with HOT+MET ($z = 1.32$; $p = .089$), but the latter takes 17.6 more min to administer, on average.

Table 2. Demographic information and neuropsychological test performance for the controls and bvFTD patients

		Control (*n* = 14)	BvFTD (*n* = 35)
Demographics	Age (years)	65.5 (6.5)	68.5 (7.2)
	Gender (M : F)	7 : 7	18 : 17
	Education (years)	13.9 (3.0)	13.6 (4.5)
Cognitive status	ACE	94.5 (5.3)	81.9 (11)
	MMSE	29.2 (1.0)	26.9 (2.9)
Classical Executive Tasks	BackDS	5.0 (1.1)	3.83 (1.4)
	Phonologic fluency	17.5 (5.7)	12.4 (7.1)
	TMT-B (sec)	94.1 (44)	182 (72)
	WCST (total score)	5.6 (0.7)	3.3 (1.8)
	WCST (pers. errors)	2.2 (2.9)	9.9 (8.0)
Executive & Social Cognition Battery	HOT (tasks completed)	4.57 (0.5)	3.26 (1.2)
	MET (rule breaks)	1.0 (1.2)	4.40 (3.0)
	IGT (net score block 5)	8.80 (8.5)	–6.51 (5.9)
	MIE (total score)	14.8 (1.4)	12.1 (1.7)
	FAUX (total score)	19.0 (1.5)	14.5 (2.5)

Note. Values are shown as *Mean* (*SD*).
ACE = Addenbrooke's Cognitive Examination; MMSE = Mini-Mental State Examination; BackDS = Backward Digit Span; TMT-B = Trail Making Test Part B; WCST = Wisconsin Card Sorting Test; HOT = Hotel Task; MET = Multiple Errands Test; IGT = Iowa Gambling Task; MIE = Mind in the Eyes; FAUX = Faux Pas.

The discriminatory accuracy of IGT+HOT was then compared with that of the longest five time-to-administer-ranked combinations (Table 5). No significant differences were found with any of these, even though the difference in mean administration time between the longer combinations and IGT+HOT ranged from 38.4 to 57.1 min.

Discriminatory accuracy was compared between IGT+HOT and measures of general cognitive status (Table 6). Accordingly, a significant difference was found between the AuC of the MMSE and IGT+HOT ($z = 1.88$; $p = .043$), and a strong trend to significance was found between the latter and the ACE ($z = 1.67$; $p = .082$). To further demonstrate the superior capacity of this abbreviated combination to differentiate bvFTD patients from controls, comparisons were made with the discriminatory capacities of classical measures of executive functioning. The AuC for IGT+HOT significantly differed from that of BackDS ($z = 2.14$; $p = .032$), PhFlu ($z = 2.70$; $p = .008$), and showed a strong trend to significance with the AuC for TMT-B ($z = 1.66$; $p = .081$), and WCST ($z = 1.57$; $p = .094$). The discriminatory accuracy of a composite score for this four-task classical executive battery

Table 3. *Mean* and *SD* Ĉ values for controls and bvFTD patients on all 31 possible combinations of tasks, sorted by descending order of AuC values

	Control (n = 14)		BvFTD (n = 35)				
Task combinations	Ĉ *Mean*	Ĉ *SD*	Ĉ *Mean*	Ĉ *SD*	AuC	*SE*	*Mean* admin time
IGT+HOT+MET+MIE	0.856	0.065	0.623	0.087	0.990	0.011	68.2
IGT+HOT+MET	0.862	0.077	0.594	0.109	0.984	0.015	63.1
IGT+HOT+MET+FAUX	0.883	0.071	0.627	0.090	0.981	0.017	84.1
IGT+HOT+MET+MIE+FAUX	0.874	0.063	0.644	0.079	0.981	0.017	89.2
IGT+MET+MIE+FAUX	0.875	0.071	0.659	0.072	0.981	0.017	70.5
IGT+MET+MIE	0.852	0.079	0.637	0.076	0.973	0.022	49.6
IGT+HOT+MIE	0.807	0.082	0.543	0.107	0.971	0.023	37.2
IGT+HOT+MIE+FAUX	0.843	0.074	0.589	0.092	0.971	0.023	58.2
HOT+MET+MIE+FAUX	0.917	0.056	0.720	0.095	0.971	0.022	75.8
IGT+MET+FAUX	0.888	0.086	0.642	0.082	0.971	0.024	65.5
IGT+MIE+FAUX	0.837	0.089	0.591	0.083	0.971	0.023	44.8
IGT+HOT+FAUX	0.843	0.090	0.549	0.110	0.971	0.024	53.1
IGT+HOT	0.790	0.112	0.460	0.149	0.963	0.028	32.1
IGT+FAUX	0.835	0.124	0.531	0.097	0.960	0.030	34.4
IGT+MIE	0.781	0.112	0.523	0.094	0.957	0.032	18.5
MET+MIE+FAUX	0.930	0.048	0.766	0.080	0.956	0.029	57.2
HOT+MET+FAUX	0.941	0.054	0.724	0.116	0.954	0.030	70.7
IGT+MET	0.860	0.110	0.600	0.101	0.952	0.033	44.4
MET+FAUX	0.971	0.038	0.794	0.099	0.952	0.030	52.0
HOT+MET+MIE	0.907	0.058	0.718	0.109	0.938	0.036	54.8
HOT+MIE+FAUX	0.902	0.072	0.673	0.116	0.937	0.038	44.8
MIE+FAUX	0.910	0.067	0.717	0.097	0.930	0.041	26.1
HOT+FAUX	0.918	0.079	0.654	0.157	0.929	0.038	39.7
FAUX	0.950	0.073	0.726	0.127	0.919	0.041	21.0
IGT	0.720	0.214	0.337	0.149	0.914	0.058	13.4
MET+MIE	0.920	0.050	0.786	0.083	0.911	0.051	36.1
HOT+MET	0.936	0.056	0.723	0.157	0.892	0.048	49.7
HOT+MIE	0.878	0.082	0.646	0.150	0.889	0.051	23.8
MIE	0.870	0.084	0.709	0.102	0.835	0.083	5.1
HOT	0.886	0.103	0.583	0.280	0.800	0.065	18.7
MET	0.993	0.023	0.863	0.139	0.784	0.069	31.0

Note. Associated standard errors (*SE*) are shown for each combination, as well as the time to administer the combination (in minutes), including instructions and scoring.
AuC = Area under the (ROC) curve.

(calculated using the procedure detailed above for the ESCB combinations) was significantly lower than that of the IGT+HOT combination ($z = 2.17$; $p = .036$).

To further investigate the generalizability of results of the IGT+HOT combination, this score was entered as the independent variable in discriminant analysis grouping controls *vs*. bvFTD patients. Box's test of equality of covariance matrices revealed that the discriminant analysis was appropriate for this dataset (Box's $M = 1.02$, $F_{1,1963.41} = .98$; $p = .33$). As expected, a significant difference was found between the groups (*Wilks' lambda* = .51; $p < .001$) on the IGT+HOT, and a canonical correlation of .87 was found between the levels of the grouping variable (controls *vs*. bvFTD) and the scores for this two-test combination (Eigenvalue = .982). LOOCV classification results revealed that 86.7% of original grouped cases and 84.4% of cross-validated grouped cases were correctly classified, revealing that the generalizability of the IGT+HOT results in this sample is very high for future independent bvFTD samples (2.3% difference between original and cross-validated).

DISCUSSION

In the present study, we investigated the usefulness of abbreviated combinations of an executive and social cognition battery (ESCB) developed by our group (Torralva et al., 2009) for the detection of subtle cognitive deficits in bvFTD patients. Because the ESCB was developed for research purposes and for clinical use, especially in testing potential bvFTD patients when diagnosis is difficult, the complete version may take over 80 min to administer and score. The main goal of this study was to propose shorter alternatives while decreasing administration times yet assuring similar high discriminatory accuracy capabilities. Our results revealed that one particular combination, represented by the IGT and the Hotel task (IGT+HOT) had similar sensitivity to

Table 4. Discriminatory accuracy comparison between IGT+HOT and (a) the five top, and (b) five bottom AuC-ranked combinations

a

AuC rank	Combination	vs. IGT+HOT	
		z	*p*
1	IGT+HOT+MET+MIE	0.88	0.37
2	IGT+HOT+MET	0.87	0.38
3	IGT+HOT+MET+FAUX	0.56	0.57
4	IGT+HOT+MET+MIE+FAUX	0.51	0.61
5	IGT+MET+MIE+FAUX	0.37	0.71

b

AuC rank	Combination	vs. IGT+HOT	
		z	*p*
31	MET	1.92	0.039
30	HOT	1.87	0.042
29	MIE	1.75	0.049
28	HOT+MIE	1.71	0.050
27	HOT+MET	1.32	0.089

AuC = Area under the (ROC) curve.

that of lengthier combinations which may take up to one more hour to administer, based on the lack of significant differences in their ability to discriminate bvFTD from controls. Moreover, this two-task combination showed high cross-validation properties revealing promising generalizability for future independent patient samples.

The use of AuC as a measure of discriminatory accuracy has been widely used in clinical research studies since the introduction of Green's well-known theorem (Green, 1964), which established that AuC equals the percentage of correct in two-alternative forced-choice scenarios. Like many other studies in all fields of medicine, the interpretation of ROC curves as measures of test accuracy (Zweig & Campbell, 1993) was used in the present study to analyze the usefulness of different task combinations of our ESCB. Comparison across the various combinations was enhanced by the mathematical transformations of the individual scores on each tasks for each combination, which was represented by a "combined mean" $\hat{C}$. Following well-established mathematical methods for the comparison of AuCs derived from the same cases, we were able to determine a discriminatory accuracy value for each combination. Briefly, the ability of the IGT+HOT combination to discriminate between bvFTD and controls did not differ significantly from that of the top-ranking combinations, all of which took much longer to administer than this two-task combination. Moreover, this combination was still more sensitive for the detection of subtle cognitive deficits in bvFTD patients than widely-used screening tools of general cognitive status. Considering the early dysexecutive syndrome that characterizes bvFTD patients (Neary et al., 1998; Hodges & Miller, 2001; Kipps, Davies, Mitchell, Kril, Halliday, & Hodges, 2007), this finding came as no surprise, as executive functions are barely tackled by the ACE, a major limitation for its use in bvFTD populations that was even acknowledged by the original authors of such tool (Mathuranath et al., 2000).

A few limitations must be taken into account in considering the generalizability of the present results. First, it could be argued that the mathematical procedure used to transform the individual scores across tests of the ESCB makes it difficult to interpret the individual results. However, a closer look at the procedure actually reveals its simplicity and utility, for it assures that performance scores are presented as a ratio of the maximum possible score. A priori analysis using simpler methods (e.g., *z*-scores relative to controls) showed some weakness due to the small sample size of our control group. Second, some criticism has been raised about the ecological validity of some of the tests used in this study. However, as shown by our previous study, the ESCB—and now also, the IGT+HOT combination—are significantly more sensitive ways to detect cognitive impairment in the early bvFTD population. Clinical experience and results of studies like ours also highlight the imperial need to develop neuropsychological tests that mimic real life scenarios, and are thus able to detect subtle cognitive deficits otherwise overlooked at by classical tests. In this sense, the administration of batteries featuring tasks such as the IGT (Bechara et al., 1994) and the Hotel Task (Manly et al., 2002) in the assessment of patients will be essential in characterizing the cognitive profile of bvFTD patients.

Naturally, future studies must be conducted to evaluate the utility of this battery and the abbreviated versions in other populations, such as psychiatric patient groups. Another important future direction is to evaluate the specificity of the ESCB and its abbreviated versions, for instance, by

Table 5. Discriminatory accuracy comparison between IGT+HOT and the five longest to administer combinations

Time to admin rank	Combination	vs. IGT+HOT		
		Time to administer difference (min)	*z*	*p*
27	IGT+MET+MIE+FAUX	38.4	0.37	0.71
28	HOT+MET+FAUX	38.6	0.17	0.86
29	HOT+MET+MIE+FAUX	43.7	0.15	0.88
30	IGT+HOT+MET+FAUX	52.0	0.56	0.57
31	IGT+HOT+MET+MIE+FAUX	57.1	0.51	0.61

Note. Administration time differences are shown.

Table 6. Discriminatory accuracy comparison between IGT+HOT and measures of general cognitive status and classical measures of executive functioning

				vs. IGT+HOT	
	Cognitive test	AuC	*SE*	*z*	*p*
General cognitive status	MMSE	0.716	0.087	1.88	0.043
	ACE	0.806	0.085	1.67	0.082
Classical tests of executive functioning	BackDS	0.757	0.095	2.14	0.032
	PhFlu	0.681	0.103	2.70	0.008
	TMT-B	0.807	0.057	1.66	0.081
	WCST	0.843	0.082	1.57	0.094
	4T-CEB	0.072	0.098	2.17	0.036

AuC = Area under the (ROC) curve; *SE* = standard error; IGT+HOT = Iowa Gambling Task + Hotel Task combination; MMSE = Mini Mental State Examination; ACE = Addenbrooke's Cognitive Examination; BackDS = Backward Digit Span; PhFlu = Phonological fluency; TMT-B = Trail Making Test Part B; WCST = Wisconsin Card Sorting Test; 4T-CEB = Four-task classical executive battery (includes the four classical tests analyzed in the present study).

analyzing the discriminatory accuracy between bvFTD and Alzheimer disease patients. Finally, the set of classical executive tasks compared against the IGT+HOT combination in this study is only limited. However, it must be acknowledged that these tasks do represent some of the most widely-used tests of executive functions worldwide.

The fact that the IGT+HOT combination was associated with the highest discriminatory accuracy in the shortest time may shed light on the nature of early impairments in bvFTD patients. While deficits are typically observed in decision-making (IGT), real-life executive demands (HOT and MET) and theory of mind (MIE and FAUX), it may be the case that the former two domains are *relatively* more impairing than theory of mind on everyday functioning. As well, tasks like MIE and FAUX involve a relatively stronger language component. Because of the substantial number of bvFTD patients who present with spared performance on standard neuropsychological batteries, it may be the case that patients benefit from their spared language abilities to perform relatively better on theory of mind tasks. This explanation would also support the dissociation between theory of mind and decision-making, which may both be impaired, but yet depend on different neural circuits within the PFC (Torralva et al., 2007). Another possibility is that, while all five tasks in the ESCB mimic real-life scenarios more closely than classical neuropsychological tests, it is the IGT and the HOT that capture everyday cognitive demands more sensitively. For instance, being able to infer what someone is thinking or feeling is undoubtedly necessary for healthy social interactions in real life. Yet, while MIE and FAUX recreate these scenarios, their assessment may be relatively less "ecological" than that of IGT and HOT. Future replications of the present findings are needed to derive stronger conclusions concerning the issue of why the IGT+HOT combination produces comparable results to the complete version of the ESCB, but the key to this matter most likely lies on both the tasks' abilities to recreate real life cognitive demands and the patients' relative impairment of each domain.

Overall, the complete version of the ESCB had been shown to be more sensitive for the detection of subtle cognitive deficits in bvFTD than standard neuropsychological tests, including classical tests of executive functioning (Torralva et al., 2009). In this study, we further demonstrated that the two-task IGT+HOT combination is more accurate in discriminating bvFTD from controls than said classical executive tests. Moreover, the IGT+HOT combination was significantly more sensitive for the detection of specific cognitive deficits in bvFTD patients than a composite score for a typical standard executive battery. This finding is crucial in the field of frontotemporal dementia, because newly emerging criteria (Rascovsky et al., 2007) from an international consortium (unpublished data) are placing a stronger focus on the neuropsychological profile of bvFTD patients. More specifically, one of the six core criteria for bvFTD stipulates that such neuropsychological profile must be characterized by deficits in executive tasks. If, as previously shown by our group (Torralva et al., 2009), there is a subset of bvFTD patients that go undetected with classical executive tests used around the globe, the availability of more sensitive tests for the detection of executive deficits will enhance early diagnosis. In this regards, the present study shows that the administration of two specific tests (approximately 30 min) of the ESCB is (a) similar in sensitivity to the administration of a more complete version of the ESCB (approximately 90 min), and (b) significantly more sensitive than the administration of a classical executive battery that may be typically found in many neuropsychological units around the world. Evidently, there are several other executive tests that are often administered to patients to assess executive function, reason why future studies should also evaluate the comparative utility of those tests in detecting executive deficits of bvFTD patients, especially during the early stages of the disease.

The present analysis attempts to suggest one shorter alternative to a solid and highly sensitive battery which had already demonstrated excellent capacity to differentiate patients with bvFTD from controls. We do not intend to replace the administration of standard cognitive tests, nor the administration of a complete version of the ESCB which is important for research purposes or for clinical settings when evaluating complex patients with potential early bvFTD. Both types of batteries provide important qualitative and quantitative information concerning the patient's neuropsychological profile. However, in clinical settings where physical (e.g., infrastructure, material, stimuli) or human (e.g., staff, time) resources are scarce, the IGT+HOT combination can be useful in the detection of specific deficits of bvFTD patients with remarkably high accuracy.

ACKNOWLEDGMENTS

We thank Nicolás Gleichgerrcht for his help in the development of mathematical and computational models for this study. This study was funded by a FINECO and a Fundación LyD grant. The authors report no conflicts of interest.

REFERENCES

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251.

Bechara, A., Damasio, A.R., Damasio, H., & Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15.

Bozeat, S., Gregory, C.A., Ralph, M.A., & Hodges, J.R. (2000). Which neuropsychiatric and behavioural features distinguish frontal and temporal variants of frontotemporal dementia from Alzheimer's disease? *Journal of Neurology, Neurosurgery, and Psychiatry*, *69*, 178–186.

Davies, R.R., Kipps, C.M., Mitchell, J., Kril, J.J., Halliday, G.M., & Hodges, J.R. (2006). Progression in frontotemporal dementia: Identifying a benign behavioral variant by magnetic resonance imaging. *Archives of Neurology*, *63*, 1627–1631.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Forman, M.S., Farmer, J., Johnson, J.K., Clark, C.M., Arnold, S.E., Coslett, H.B., et al. (2006). Frontotemporal dementia: Clinicopathological correlations. *Annals of Neurology*, *59*, 952–962.

Green, D.M. (1964). General prediction relating yes-no and forced-choice results. *The Journal of the Acoustical Society of America*, *36*, 1042.

Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*, 839–843.

Hodges, J.R., & Miller, B. (2001). The neuropsychology of frontal variant frontotemporal dementia and semantic dementia. Introduction to the special topic papers: Part II. *Neurocase*, *7*, 113.

Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A., & Martin, R.L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, *140*, 566–572.

Josephs, K.A. (2008). Frontotemporal dementia and related disorders: Deciphering the enigma. *Annals of Neurology*, *64*, 4–14.

Kipps, C.M., Davies, R.R., Mitchell, J., Kril, J.J., Halliday, G.M., & Hodges, J.R. (2007). Clinical significance of lobar atrophy in frontotemporal dementia: Application of an MRI visual rating scale. *Dementia and Geriatric Cognitive Disorders*, *23*, 334–342.

Kipps, C.M., Knibb, J.A., Patterson, K., & Hodges, J.R. (2008). Chapter 28 Neuropsychology of frontotemporal dementia. *Handbook of Clinical Neurology 88*, 527–548.

Kipps, C.M., Nestor, P.J., Fryer, T.D., & Hodges, J.R. (2007). Behavioural variant frontotemporal dementia: Not all it seems? *Neurocase*, *13*, 237–247.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological assessment*. New York: Oxford University Press.

Manly, T., Hawkins, K., Evans, J., Woldt, K., & Robertson, I.H. (2002). Rehabilitation of executive function: Facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, *40*, 271–281.

Mathuranath, P.S., Nestor, P.J., Berrios, G.E., Rakowicz, W., & Hodges, J.R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*, *55*, 1613–1620.

Mendez, M.F., Shapira, J.S., McMurtray, A., Licht, E., & Miller, B.L. (2007). Accuracy of the clinical evaluation for frontotemporal dementia. *Archives of Neurology*, *64*, 830–835.

Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., et al. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*, 1546–1554.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*, 313–324.

Partington, J.E., & Leiter, R. (1949). Partington's pathway test. *The Psychological Service Center Bulletin*, *1*, 9–20.

Rascovsky, K., Hodges, J.R., Kipps, C.M., Johnson, J.K., Seeley, W.W., Mendez, M.F., et al. (2007). Diagnostic criteria for the behavioral variant of frontotemporal dementia (bvFTD): Current limitations and future directions. *Alzheimer Disease and Associated Disorders*, *21*, S14–S18.

Shallice, T., & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(Pt 2), 727–741.

Stone, V.E., Baron-Cohen, S., & Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640–656.

Torralva, T., Kipps, C.M., Hodges, J.R., Clark, L., Bekinschtein, T., Roca, M., et al. (2007). The relationship between affective decision-making and theory of mind in the frontal variant of frontotemporal dementia. *Neuropsychologia*, *45*, 342–349.

Torralva, T., Roca, M., Gleichgerrcht, E., Bekinschtein, T., & Manes, F. (2009). A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain*, *132*(Pt 5), 1299–1309.

Wechsler, D. (1997). *Wechsler memory scale, complete kit* (3rd ed.). San Antonio, TX: The Psychological Corporation, 1997.

Zweig, M.H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, *39*, 561–577.