

L'uso degli strumenti neuropsicologici nell'ambito del Progetto Cronos

Goffredo BIANCHI (a), Marina GASPARINI (b), Bruno CAFFARI (a),
Giacoma C. SORRENTINO (a), Clara BIANCHI (a), Giuseppe BRUNO (b),
Marina MAGGINI (a), Roberto RASCHETTI (a) e Nicola VANACORE (a)

(a) Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute,
Istituto Superiore di Sanità, Roma

(b) Dipartimento di Scienze Neurologiche, Università degli Studi "La Sapienza", Roma

Riassunto. - Nell'ambito del Progetto Cronos sono state attivate sull'intero territorio nazionale 503 Unità di Valutazione Alzheimer (UVA). Nel periodo di giugno/luglio 2002 è stata inviata a tutti i centri UVA una scheda di rilevazione tramite la quale veniva raccolta l'informazione sugli strumenti neuropsicologici comunemente utilizzati nel processo diagnostico delle demenze. Solo 196 delle UVA che hanno risposto al questionario (196/392) dichiarano di utilizzare test e batterie neuropsicologiche, mentre il 97,8% dichiara di utilizzare il *mini mental state examination* (MMSE). Questo contributo ha l'obiettivo di censire, per la prima volta in Italia, l'uso degli strumenti neuropsicologici comunemente utilizzati nelle strutture assistenziali deputate alla diagnosi e al trattamento delle demenze e di proporre un approfondimento metodologico strettamente legato alla validazione dei test.

Parole chiave: test neuropsicologici, Progetto Cronos, demenza di Alzheimer.

Summary (*Use of neuropsychological tests in the Cronos Project*). - Within the Cronos Project, 503 Alzheimer's disease units were activated throughout the country. In June/July 2002 a questionnaire was sent to all Alzheimer's disease units to collect data on neuropsychological tests utilised in the diagnostic process of dementias. Only 196 of Alzheimer's disease units that have responded to the questionnaire (196/392) declare to utilize neuropsychological test while 97.8% use the mini mental state examination. This paper aims to collect, for the first time in Italy, data on use of the neuropsychological tests in the specialistic units on dementia and to propose a discussion on the reliability of neuropsychological instruments

Key words: neuropsychological tests, Cronos Project, Alzheimer's disease.

Introduzione

Negli ultimi anni, in concomitanza con l'immissione sul mercato degli inibitori della colinesterasi, si è assistito ad una maggiore attenzione della comunità scientifica verso le procedure diagnostiche utilizzate nella pratica clinica per porre la diagnosi di demenza. Ciò ha comportato una notevole diffusione dell'uso degli strumenti neuropsicologici che, inizialmente elaborati in un ambito di ricerca, sono stati poi utilizzati nella routine diagnostica senza una adeguata valutazione critica. La diffusione delle pratiche neuropsicologiche ha quindi indotto qualche difficoltà nel gestire la tipologia e la mole di dati elaborati da un laboratorio di neuropsicologia. Contemporaneamente, vista la crescente rilevanza in termini di sanità pubblica del problema delle demenze, l'attenzione della comunità scientifica si è indirizzata verso la possibile identificazione delle fasi precoci della demenza; in particolare

si è tentato di caratterizzare quella situazione, definita come *mild cognitive impairment*, nella quale è presente un deficit cognitivo isolato. Questa categoria, qualora si riuscisse a ben caratterizzarla, potrebbe costituire per un sottogruppo di soggetti un fattore di rischio per l'insorgenza della demenza. Il ruolo degli strumenti neuropsicologici ha assunto in questo ambito un'importanza per alcuni aspetti ancor maggiore rispetto all'uso che se ne fa per porre la diagnosi di demenza, in quanto tutte le questioni metodologiche relative alla validità o meno del singolo test neuropsicologico si amplificano, e riportano all'attenzione degli esperti il problema del corretto uso di tali strumenti [1]. In Italia è stato condotto il più ampio studio osservazionale post-marketing, all'interno del Progetto Cronos, sull'uso degli inibitori della colinesterasi nella pratica clinica corrente della demenza di Alzheimer. Tale studio ha indotto l'attivazione di 503 Unità di Valutazione per l'Alzheimer (UVA) diffuse sull'intero

territorio nazionale [2]. Questo contributo ha l'obiettivo di censire, per la prima volta in Italia, l'uso degli strumenti neuropsicologici comunemente utilizzati nelle strutture assistenziali deputate alla diagnosi ed al trattamento delle demenze (UVA) e di proporre un approfondimento metodologico strettamente legato alla validazione dei test.

In questo contesto la predittività degli strumenti neuropsicologici rappresenta un tema rilevante. Un test si definisce attendibile (affidabile, o anche fedele) quando i punteggi ottenuti da un gruppo di soggetti sono coerenti, stabili nel tempo e costanti dopo molte somministrazioni e in assenza di cambiamenti evidenti quali variazioni psicologiche e fisiche degli individui che si sottopongono al test, o anche dell'ambiente in cui questo ha luogo [3]. La grande importanza del coefficiente di attendibilità sta nell'esprimere il grado di fiducia che è lecito riporre in un test. Uno strumento è tanto più attendibile quanto più è libero da errori di misura, in modo tale che i punteggi con esso ottenuti corrispondano il più possibile ai punteggi veri. Tutto ciò può essere concettualmente espresso dalla seguente equazione:

$$V_x = V_v + V_e$$

dove la varianza dei punteggi ottenuti (V_x) può essere considerata come la somma della varianza dei punteggi veri (V_v) e della varianza dovuta agli errori di misura (V_e). L'attendibilità di un test esprime in altre parole la misura in cui le differenze fra i punteggi di un gruppo di soggetti sottoposti allo stesso test in tempi diversi, o a versioni equivalenti dello stesso, possano essere attribuite a errori casuali nella misurazione o all'effettivo variare nei soggetti delle caratteristiche misurate. Quando si parla di un test, quindi, un livello accettabile di affidabilità è qualcosa di più importante che una semplice questione accademica; diversi autori suggeriscono che il coefficiente di attendibilità ($r_{xx} = V_v/V_x$) dovrebbe essere almeno di 0,90 quando un test viene utilizzato per prendere decisioni riguardanti gli individui [4].

Grande importanza assumono in questo quadro i due indicatori psicometrici più utilizzati per valutare l'attendibilità di un test:

- *test-retest reliability*: valuta il grado di correlazione esistente tra due distribuzioni di punteggi ottenuti dagli stessi soggetti nello stesso test somministrato loro in tempi diversi;

- *inter-rater reliability*: valuta il grado di correlazione esistente tra due o più distribuzioni di punteggi assegnati da altrettanti correttori, indipendentemente l'uno dall'altro, ad un campione di protocolli ottenuti dagli stessi soggetti nello stesso test.

La *test-retest reliability* rileva soprattutto la dimensione della stabilità dei punteggi nel tempo, ed è calcolata in genere con il coefficiente di correlazione di Pearson. Se il test risulta perfettamente attendibile significa che i soggetti hanno ottenuto lo stesso tipo di

punteggio in entrambe le somministrazioni, cioè il punteggio nella seconda prova potrà essere previsto a partire dal primo.

L'*inter-rater reliability*, o affidabilità tra scorers, misura la varianza d'errore legata ai criteri di attribuzione dei punteggi ed alla soggettività dei valutatori. La procedura è abbastanza semplice: due correttori assegnano punteggi ad un campione di protocolli indipendentemente l'uno dall'altro; successivamente tali punteggi si correlano accoppiandoli. Più alto è il coefficiente di correlazione ottenuto, minore è la varianza d'errore dovuta a chi attribuisce i punteggi. Avremo così che un coefficiente di 0,78 indicherà che il 78% della variazione dei punteggi nel test corrisponde alla varianza vera, dovuta cioè alle differenze individuali, mentre il 22% rappresenterà la variazione dei punteggi per cause non controllabili (varianza d'errore).

Poiché le prestazioni individuali ai test assumono un significato e sono interpretabili solo se confrontate con specifici criteri (prestazioni medie di gruppi di soggetti o performance tipiche), tali prestazioni devono essere ottenute in condizioni standardizzate, cioè uniformi. Il concetto di standardizzazione, oltre a significare uniformità nella procedura di somministrazione del test e di valutazione delle risposte al fine di comparare i risultati individuali, sta ad indicare la determinazione delle "norme" statistiche di un test. Possiamo a questo punto introdurre la definizione di campione così come viene comunemente presentata: "Standardizzare o tarare un test significa somministrarlo ad un ampio gruppo di soggetti statisticamente rappresentativo della popolazione per la quale il test verrà usato. Tale gruppo, chiamato campione di standardizzazione serve per stabilire le *norme* del test, cioè le prestazioni, che per la popolazione da cui il campione è stato casualmente estratto, sono da considerarsi *normali* cioè *medie*" [3]. Senza queste norme prestazionali non è quindi possibile interpretare i punteggi individuali, cioè attribuire loro un significato e un valore. Il punteggio grezzo infatti di per sé non ha alcun significato se non confrontato con dei valori normativi. Questi ultimi sono costituiti non solo dalle prestazioni medie del campione di taratura, ma anche dalla deviazione standard, ossia dal grado di scostamento dei punteggi al di sopra o al di sotto di tale media. Grazie a questi parametri di riferimento è possibile conoscere quale è la posizione occupata da un individuo all'interno del campione normativo e stabilire se il suo punteggio può essere considerato normale, e di quanto superiore o inferiore alla media. Il numero di soggetti utilizzati come gruppo normativo o di standardizzazione, quindi, deve essere il più ampio possibile, anche se la dimensione di per sé non è una garanzia della sua rappresentatività. Perché sia tale deve essere accuratamente selezionato secondo rigorose procedure statistiche di campionamento.

Rilevazione degli strumenti neuropsicologici utilizzati nell'ambito del Progetto Cronos

Nel periodo di giugno/luglio 2002 è stata inviata a tutti i centri UVA attivi nell'ambito del Progetto Cronos una scheda di rilevazione. I dati raccolti attraverso questo strumento hanno consentito di descrivere alcune caratteristiche strutturali e alcune modalità di funzionamento delle UVA stesse. È stato quindi possibile disegnare un quadro delle diverse localizzazioni delle unità, delle disponibilità di personale, delle strumentazioni e delle diverse metodologie utilizzate per le diagnosi.

Trecentonovantadue su 503 (77,9%) UVA che hanno risposto al questionario impiegano complessivamente oltre 2000 persone tra medici specialisti (neurologi, psichiatri e geriatri), medici specializzandi, psicologi, infermieri, terapisti della riabilitazione, assistenti sociali e tecnico amministrativi. Soltanto 196 centri (50%) dichiarano di aver utilizzato test neuropsicologici/batterie neuropsicologiche per la diagnosi di demenza di Alzheimer mentre le scale cliniche *mini mental state examination* (MMSE), *instrumental activities of daily living* (IADL) e *activities of daily living* (ADL), così come richiesto dal protocollo del Progetto Cronos, vengono impiegate nella stragrande maggioranza dei centri (97,8%).

L'indagine ha permesso di evidenziare una tipologia di strumenti di valutazione neuropsicologica utilizzati nei centri UVA molto variegata. Il dato che emerge è poco confortante: 46 singoli test, 8 batterie, 24 scale cliniche, 23 test non strutturati e 49 test non identificati. Nella Tab. 1 sono riportati i principali test e batterie neuropsicologiche con le percentuali dei centri UVA che dichiarano di utilizzarli.

Nella Tab. 2 sono riportati i test e le batterie neuropsicologiche per le quali è stato possibile individuare la numerosità del campione di riferimento della popolazione italiana ed i parametri che esprimono la variabilità del fenomeno (*test-retest e inter-rater reliability*).

Nelle Tab. 3-4 sono invece riassunte le caratteristiche dei campioni normativo della popolazione italiana maggiormente utilizzati in ambito neuropsicologico per sesso, età, scolarità e tipologia di campione [5-14].

Discussione

Il quadro generale di come viene eseguita la valutazione cognitiva in Italia sui pazienti con possibile demenza di Alzheimer appare preoccupante. Da un lato il 50% delle UVA dichiara di non effettuare alcuna valutazione neuropsicologica, presentando probabilmente alcune difficoltà nell'identificazione

del deficit cognitivo e nella diagnosi differenziale delle demenze. In questa situazione diviene problematico un raccordo scientifico e clinico fra le diverse UVA.

Si ricorda infatti che il MMSE, l'unico strumento di valutazione dello stato mentale ampiamente utilizzato, ha una bassa sensibilità (capacità di individuare come veri i soggetti con disturbo cognitivo) ed è stato costruito per essere adoperato in una fase di screening non nella misurazione del deficit cognitivo. Dall'altro lato i 25 test neuropsicologici più utilizzati difficilmente raggiungono un elevato standard psicometrico: 6 di essi (pari al 24%) non hanno indici psicometrici, 14 (56%) hanno una valutazione *test-retest* basata probabilmente sul coefficiente di Pearson, ed il livello di attendibilità solo in 3 casi è eguale al valore di 0,90 che viene considerato da alcuni autori il livello minimo accettabile

Tabella 1. - Principali strumenti di valutazione neuropsicologica adottati dalle Unità di Valutazione Alzheimer (n. = 392)

Strumenti di valutazione	Unità di Valutazione Alzheimer	
	(n.)	(%)
Test		
Matrici attenzionali	72	18,4
Fluenza verbale alfabetica	67	17,1
Fluenza verbale per categorie	60	15,3
Token	52	13,3
Corsi	47	12,0
Memoria di prosa	44	11,2
15 parole di Rey	43	11,0
Orologio	38	9,7
Matrici progressive di Raven 1938	24	6,1
Figura complessa di Rey	22	5,6
Digit Span	19	4,8
Ripetizione seriale di parole bisillabiche	9	2,3
Stroop	9	2,3
Boston	7	1,8
Aachener aphasia	6	1,5
Matrici progressive standard (MPS)	6	1,5
Weigl's sorting test	6	1,5
Matrici progressive colori	5	1,3
Memoria visiva immediata	5	1,3
Trail making test	5	1,3
Associated learning test (ALT)	4	1,0
Prove di apprendimento motorio	2	0,5
Discriminazione sgorbi	1	0,3
Giudizi verbali	1	0,3
Wisconsin carding sorting test	1	0,3
Batterie		
Alzheimer's disease assessment scale-cog	94	24,0
Milan overall dementia assessment	92	23,5
Severe impairment battery (SIB)	13	3,3
Wechsler adult intelligence scale revised	12	3,1
Wechsler memory scale (WMS)	6	1,5
Spinnler e Tognoni (SET)	5	1,3

Tabella 2. - Parametri psicometrici relativi ai campioni normativi italiani

Test e batterie	Campione popolazione italiana (n.)	Campione test-retest (n.)	Test-retest reliability	Inter-rater reliability
Matrici attenzionali	321	35	0,53	
Fluenza verbale alfabetica	320	n.d.	0,75-0,84	
Fluenza verbale categorie	321	35	0,92	
Token	321	35	0,67	
Corsi	321	35	0,38	
Memoria di prosa	321	35	0,69	
Matrici progressive Raven 1938	321	35	0,89	
Figura complessa di Rey	227	n.d.		0,95-0,98
Digit span	1355	n.d.		
Stroop test	209	n.d.	0,80-0,88	
Weigl	321	35		
Matrici progressive colori	894	n.d.		
Trail making test	287	57	0,74-0,85	
Discriminazione sgorbi	321	35	0,75	
Span verbale	321	35	0,71	
Giudizi verbali	321	35	0,43	
Wisconsin card sorting test	205	n.d.		
Alzheimer disease assessment scale (ADAS) cog	69	69	0,90-0,94	0,99
Milan overall dementia assessment (MODA)	217	n.d.	0,83	
Wechsler adult intelligence scale (WAIS)	1630	n.d.	0,81-0,91	
Mental deterioration battery	340	n.d.		

n.d.: non disponibile.

Tabella 3. - Caratteristiche socio-anagrafiche dei campioni normativi della popolazione italiana (test neuropsicologici)

Variabili	Spinnler-Tognoni [5]	Orsini et al. [6]	Measso et al. [7]	Giovagnoli et al. [8]	Barbarotto et al. [9]	Carlesimo et al. [10]	Caffarra et al. [11]	Caffarra et al. [12]
N. soggetti	321	1355	894	287	209	227	280	248
≥ 60 aa.	214	n.d.	197	49	69	101	120	97
M/F	0,72	0,87	0,95	0,86	0,71	0,94	1	1
Scolarità (aa) (M± DS)	-	9,44 ± 5,52	8,39 ± 4,51	11,4 ± 4,7	10,8 ± 4,8	12,4	11,1 ± 4,81	11,54 ± 4,75
Scolarità (livello)								
Elementare	147 (45,8%)	-	362 (40,5%)	63 (21,9%)	53 (25,3%)	-	50 (17,9%)	33 (13,3%)
Media inferiore	65 (20,2%)	-	192 (21,4%)	56 (19,6%)	-	-	74 (26,5%)	68 (27,5%)
Media superiore	72 (22,4%)	-	251 (28,2%)	114 (39,7%)	61 (29,2%)	-	78 (27,8%)	71 (28,6%)
Laurea	37 (11,6%)	-	89 (9,9%)	54 (18,8%)	95 (45,5%)	-	78 (27,9%)	76 (30,6%)
Tipologia del campione	Congiunti di pazienti ambulatoriali	n.d.	Estratto da liste residenti di 6 località	n.d.	n.d.	Soggetti ospedalizzati	Volontari	Volontari

n.d.: non disponibile.

per l'applicazione dei test a livello individuale; l'*inter-rater-reliability* è riportato solo per il 20% dei test ed in nessun caso è stato utilizzato il parametro richiesto dagli standard di qualità quale l'indice kappa [15].

Per quanto riguarda le 6 batterie neuropsicologiche più utilizzate, in 5 casi si dispone di un *test-retest reliability*, in 2 casi di un *inter-rater reliability* ed in 1 non vi sono indici psicometrici.

Inoltre i test non strutturati e quelli non identificati vengono utilizzati dal 5-7% delle UVA.

Un'ulteriore considerazione riguarda la natura stessa degli strumenti, molti dei quali segnalati in modo ambiguo, non strutturati, di vecchia concezione e quindi caduti in disuso da anni perché sostituiti da strumenti più sensibili, o inadeguati ad indagare i deficit cognitivi dell'Alzheimer. Inoltre da questa indagine è

Tabella 4. - Caratteristiche socio-anagrafiche dei campioni normativi della popolazione italiana (batterie neuropsicologiche)

Variabili	Brazzelli <i>et al.</i> [13]	Carlesimo <i>et al.</i> [14]
N. soggetti	217	340
60 aa.	126	144
M/F	0,90	1,1
Scolarità (aa) (M± DS)	9,1 ± 4,9	10,2 ± 4,3
Tipologia del campione	n.d	Soggetti ospedalizzati

emerso che in molte UVA vengono utilizzati test o batterie standardizzati all'estero perché non ancora standardizzati in Italia, compromettendo quindi la rappresentatività del campione e di conseguenza l'affidabilità dello strumento neuropsicologico.

Un punto di debolezza di molti di questi strumenti è quello dell'attendibilità, cioè il grado di accuratezza o di precisione di una procedura di misurazione: i coefficienti variano dallo 0,26 allo 0,90, e i test sono spesso privi di specifiche validazioni per la popolazione italiana. Gli studi di taratura effettuati in Italia, inoltre, sono spesso caratterizzati da campioni poco numerosi e composti da fasce socio-economiche ristrette, col rischio di essere scarsamente rappresentativi della popolazione generale. Per giunta le operazioni test-retest vengono in genere effettuate su una percentuale ridotta dello stesso campione per motivi di natura economica. Le conseguenze sul piano clinico sono importanti, soprattutto per quanto riguarda il valore predittivo degli strumenti.

Per quanto riguarda le batterie è da segnalare l'uso di strumenti nati nell'ambito dei protocolli di sperimentazione farmacologica, e il cui scopo non è tanto diagnostico-cognitivo quanto quello di monitoraggio "prima-durante-dopo" il trattamento.

Questa situazione pone al centro della riflessione due questioni di carattere generale: quali test utilizzare nella pratica neuropsicologica corrente? Quale valore dare nel percorso diagnostico delle demenze al risultato del test neuropsicologico?

Nel primo caso non si possono non privilegiare quei test validati sulla popolazione italiana e che raggiungono un livello accettabile dal punto di vista psicométrico, mentre per quanto riguarda il secondo aspetto bisogna che il medico abbia una elevata competenza per inserire il valore numerico ed il punto di vista del neuropsicologo nella storia clinica del soggetto. Il livello purtroppo non eccelso degli strumenti neuropsicologici disponibili sottolinea l'esigenza di effettuare una nuova validazione degli

stessi test su un campione realmente rappresentativo della popolazione italiana con un approccio psicométrico moderno e rigoroso.

In conclusione, alla luce di quanto emerso dall'indagine sui 392 centri UVA e di quanto precedentemente detto riguardo la costruzione e l'utilizzo dei test, possiamo affermare che:

- le UVA che non utilizzano test o batterie neuropsicologiche presentano probabilmente alcune difficoltà nell'identificazione del deficit cognitivo e nella diagnosi differenziale delle demenze;

- l'affidabilità dei test neuropsicologici utilizzati valutata secondo i parametri della psicomètria moderna, appare modesta;

- le caratteristiche socio-anagrafiche e la tipologia dei campioni sui quali è stato validato l'uso dei test e delle batterie neuropsicologiche in Italia mostrano alcune lacune in termini di rappresentatività. Infatti i meccanismi di validazione nella popolazione generale e i processi di elaborazione di un punteggio "normale" della funzione cognitiva appaiono inadeguati in relazione alla crescente rilevanza in termini di sanità pubblica dell'uso dei test neuropsicologici;

- l'identificazione di una variazione di punteggio, clinicamente rilevante, in una scala clinica o in un test neuropsicologico, dovrebbe basarsi su una accurata valutazione psicométrica prima di essere adottata in un clinical trial o nella pratica clinica corrente.

Lavoro presentato su invito.
Accettato il 17 ottobre 2004.

BIBLIOGRAFIA

1. Levin HS. A guide to clinical neuropsychological testing. *Arch Neurol* 1994;51(9):854-59.
2. Raschetti R, Maggini M, Sorrentino GC, Martini N, Caffari B, Vanacore N. A cohort study of effectiveness of acetylcholinesterase inhibitors in Alzheimer's Disease. *Eur J Clin Pharmacol* (in stampa).
3. Pedrabissi L, Santinello M. *I test psicologici*. Bologna: Il Mulino; 1997.
4. Salvia J, Ysseldyke S. *Assessment in special and remedial education*. Boston: Mass. Houghton Mifflin; 1988.
5. Spinnler H, Tognoni G. Standardizzazione e taratura italiana di test neuropsicologici. *Ital J Neurol Sci* 1987;Suppl 8(6).
6. Orsini A, Grossi D, Capitani E, Laiacona M, Papagno C, Vallar G. Verbal and spatial immediate memory span. Normative data from 1355 adults and 1112 children. *Ital J Neurol Sci* 1987;8(6):539-48.
7. Measso G, Zappalà G, Cavarzeran F, Crook TH, Romani L, Pirozzolo FJ, Grigoletto F, Amaducci LA, Massari D, Lebowitz BD. Raven's colored progressive matrices: a normative study of a random sample of healthy adults. *Acta Neurol Scand* 1993;88(1):70-4.

8. Giovagnoli AR, Del Pesce M, Mascheroni S, Simoncelli M, Laiacona M, Capitani E. Trail Making Test: normative values from 287 normal adult controls. *Ital J Neurol Sci* 1996;17(4):305-9.
9. Barbarotto R, Laiacona M, Frosio R, Vecchio M, Farinato A, Capitani E. A normative study on visual reaction times and two Stroop colour-word tests. *Ital J Neurol Sci* 1998;19(3):161-70.
10. Carlesimo GA, Buccione I, Fadda L, Graceffa A, Mauri M, Lorusso S, Bevilacqua G, Caltagirone C. Standardizzazione di due test di memoria per uso clinico. Breve racconto e figura di Rey. *Nuova Rivista di Neurologia* 2002;12(1):1-13.
11. Caffarra P, Vezzadini G, Dieci F, Zonato F, Venneri A. Rey-Osterrieth complex figure: normative values in an Italian population sample. *Neurol Sci* 2001;22(6):443-7.
12. Caffarra P, Vezzadini G, Dieci F, Zonato F, Venneri A. Modified Card Sorting Test: normative data. *J Clin Exp Neuropsychol* 2004;26(2):246-50.
13. Brazzelli M, Capitani E, Della Sala S, Spinnler H, Zuffi M. *MODA. Milan Overall Dementia Assessment. Manuale*. Firenze: Organizzazioni Speciali; 1994.
14. Carlesimo GA, Caltagirone C, Gainotti G. The mental deterioration battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. *Eur Neurol* 1996;36(6):378-84.
15. Sharrack B, Hughes RAC, Soudain S, Dunn G. The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain* 1999;122:141-59.