Normative Study of the Italian Language TOMMORROW Neurocognitive Battery

Brenda L. Plassman[1,2]; Samantha Galluzzi[3]; Carlo de Lena[4]; Kathleen A. Welsh-Bohmer[1,2]; Heather Romero[1,2]; Kathleen M. Hayden[1,2,5]; Kumar Budur[6]; Grant Runyan[6]; Janet O'Neil[6]; Stephen Crawford[7]; Mark Atkinson[7]; Toyoko Oguri[7]; Alexandra Atkins[8]; Nicole Turcotte[8]; Richard Keefe[2,8]; Shyama Brewster[9]; Dominic Fitzsimmons[10]; Daniel K. Burns[9]; Maura Parapini[3,11]; Letizia Imbriano[4]; Giovanni B. Frisoni[3,11]; and Allen D. Roses[9] for the TOMMORROW study investigators

[1]Joseph and Kathleen Bryan ADRC, Duke University Medical Center, Durham, NC, USA; [2]Department of Psychiatry, Duke University Medical Center, Durham, NC, USA; [3]IRCCS Fatebenefratelli, Brescia, Italy; [4]Department of Neurology and Psychiatry, Sapienza University of Rome, Italy; [5]Department of Social Sciences and Health Policy, Wake Forest School of Medicine, Winston-Salem, NC, USA; [6]Takeda Development Center Americas, Inc., Deerfield, IL, USA; [7]Covance Inc., Princeton, NJ, USA; [8]NeuroCog Trials, Durham, NC, USA; [9]Zinfandel Pharmaceuticals, Inc., Chapel Hill, NC, USA; [10]Takeda Development Centre Europe Ltd, London, UK; [11]University Hospitals and University of Geneva, Geneva, Switzerland

Introduction

Global clinical trials are increasingly focused on the presymptomatic phase of Alzheimer's disease (AD) with the goal of delaying the onset of AD symptoms. These trials require a comprehensive neurocognitive battery capable of detecting subtle cognitive change as individuals transition from normal cognition to mild cognitive impairment due to AD (MCI-AD).

However, few neuropsychological tests have been validated or normed for use in non-English-speaking countries. For use in global clinical trials, neuropsychological tests need to be shown to be valid, reliable, and culturally sound in different languages.

The TOMMORROW Study is an international trial using MCI-AD as the primary endpoint. In order to facilitate assessment of participants expected to enroll in future clinical trials in Italian-speaking countries, we conducted a validation and normative study of the TOMMORROW neurocognitive battery in Rome and Brescia, Italy.

Methods

Linguistic and cultural adaptation of the neuropsychological battery was completed in Italian. This was done in accordance with International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines, including forward and backward translation, cognitive debriefing interviews, in-country expert review, and review by neuropsychologists.

Sample

Cognitively Normal Controls

- Approximately n=50 in each of 4 age strata (65-69, 70-74, 75-79, 80-88).
- Each age strata required to have a minimum of 15 males and 15 females. Each strata also required to have between 5 and 10 individuals with low education, defined as <8 years of schooling.
- Required to have Mini Mental State Examination (MMSE) scores >24 and complete clinical evaluation to confirm normal cognition.
- Mean standard deviation (SD) MMSE score = 28.9 (1.31); range (25-31) with age and education adjustment.

Individuals with AD

- N=25 with clinically confirmed AD based on National Institute of Neurological Communicative Disorders and Stroke and the Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) criteria.
- Mean (SD) age = 76.0 (5.57); range (67-86).
- Mean (SD) MMSE score = 19.2 (3.42); range (14-27) indicating mild-moderate dementia.

Results

Criterion Validity

- Criterion validity was demonstrated by greater impairment shown in the AD group on all tests compared with controls (**Table 1**). Visual attention (Trails A) had the strongest effect size. This may be due to large differences in education between normal controls and AD subjects. In addition, the discontinuation time limit for Trails A was 300 seconds as stated in the test manual, rather than the 180-second limit that is more commonly used in research and clinical practice. This may have accentuated the differences between the 2 groups.
- **Table 2** shows the performance of the California Verbal Learning Test – 2nd edition (CVLT-II) Delay Recall in discriminating between AD group and controls.
- Composite scores derived from 4 cognitive domains (episodic memory, executive function, attention, and language) did not improve the discrimination between the AD group and controls (**Table 2**).

Construct Validity

- Supported by moderate to high correlations among tests in related domains (0.53-0.90).

Test-Retest and Alternate Form Reliability

- Supported by adequate correlations, most >0.60 (**Table 3**).

Table 1. Criterion Validity – Comparison of AD and Cognitively Normal Controls

Test	NC (n=188)	AD (n=25)	F-Statistic	Cohen's *d*
Trails A	52.3 (2.7)	158.0 (7.7)	65.82*	2.97
CVLT-II Long-Delay Free Recall Correct	8.5 (0.2)	1.3 (0.6)	69.77*	2.49
BVMT-R Delayed Recall	7.4 (0.2)	1.3 (0.5)	60.73*	2.44
Trails B	132.6 (4.1)	257.5 (12.4)	65.06*	2.30
MINT	27.8 (0.2)	20.5 (0.7)	52.87*	2.25
CVLT-II Short-Delay Free Recall Correct	7.8 (0.2)	1.6 (0.6)	55.81*	2.13
Clock Drawing Test	8.1 (0.1)	5.1 (0.3)	30.15*	1.97
Semantic Fluency (animals)	18.4 (0.3)	11.0 (0.9)	38.00*	1.77
BVMT-R Copy Accuracy	11.0 (0.1)	8.1 (0.4)	19.01*	1.69
Digit Span Backward	5.4 (0.1)	3.1 (0.3)	38.32*	1.52
Lexical Fluency	32.2 (0.7)	19.7 (1.9)	31.26*	1.43
Digit Span Total	13.9 (0.2)	10.3 (0.6)	33.33*	1.26
Digit Span Forward	8.4 (0.1)	7.30 (0.4)	14.91*	0.66

Notes: *P<0.001. Analyses adjusted for age, years of education, and gender.
BVMT-R: Brief Visual Memory Test–Revised; CVLT-II: California Verbal Learning Test – 2nd edition; MINT: Multilingual Naming Test; NC: normal controls.

Table 2. Criterion Validity – Discrimination of AD Cases from Normal Controls

Model	Sensitivity	PPV	Specificity	NPV	Pseudo R^2	Age	Gender	Education
CVLT-II Long-Delay Free Recall	92.0	85.2	97.9	98.9	0.406	—	—	—
CVLT-II Long-Delay Free Recall w/Covariates	96.0	96.0	99.5	99.5	0.442	0.027 (P=.01)	0.001 (ns)	0.005 (ns)
CVLT-II Short-Delay Free Recall	88.0	81.5	97.3	98.4	0.387	—	—	—
CVLT-II Short-Delay Free Recall w/Covariates	84.0	84.0	97.9	97.9	0.409	0.004 (ns)	0.008 (ns)	0.009 (ns)
Composite Score	79.2	79.2	97.3	97.3	0.39	—	—	—
Composite Score w/Covariates	87.5	87.5	98.4	98.4	0.43	0.033 (P=0.008)	0.008 (ns)	0.017 (P=0.033)

CVLT-II: California Verbal Learning Test – 2nd edition; NPV: negative predictive value; ns: not significant; PPV: positive predictive value.

Table 3. Test-Retest and Alternate Form Reliability

Test	Test-Retest Reliability		Alternate Form Reliability	
	N	Pearson *r*	N	Pearson *r*
CVLT-II Short-Delay Free Recall	108	0.800*	72	0.635*
CVLT-II Long-Delay Free Recall	107	0.830*	72	0.716*
BVMT-R Delay Recall	106	0.783*	72	0.650*
Digit Span Forward	180	0.645*	—	—
Digit Span Backward	179	0.680*	—	—
Digit Span Total	179	0.774*	—	—
Trails A Total Seconds	178	0.791*	—	—
Trails B Total Seconds	176	0.818*	—	—
MINT Total	180	0.861*	—	—
Semantic Fluency	180	0.706*	—	—
Lexical Fluency: Total Words	180	0.845*	—	—
BVMT-R Copy†	106	0.356*	—	—
Clock Drawing Test†	178	0.514*	—	—

Note: Standard forms for CVLT-II or Form 1 for BVMT-R.
†Spearman rank used in place of Pearson for highly skewed distributions.
*P<.001. BVMT-R: Brief Visual Memory Test–Revised; CVLT-II: California Verbal Learning Test – 2nd edition; MINT: Multilingual Naming Test.

Developing Norms

- As expected, for many tests, older age and, to a lesser extent, less education were associated with poorer cognitive test performance compared with younger age and greater education. Gender had a limited effect on performance.
- These results support the use of age-stratified norms. Given the relatively small number of individuals with low education within each age stratum, education adjustment was not warranted.

Conclusions

- The Italian translation and cultural adaptation of the TOMMORROW neurocognitive battery is psychometrically sound and generally performs comparably to English test versions.
- However, the variability in performance on some tests observed between the English version in the US and the Italian version highlights the need for region-specific norms.
- Based on these findings, which demonstrate adequate reliability and validity, age-corrected normative values were developed for use by Italian-speaking older adults.
- The normative data derived from this validation study is a tool for valid interpretation of cognitive performance in Italian-speaking individuals and for identifying cases of MCI-AD in clinical trials.
- To date, few global clinical trials have empirically evaluated the performance characteristics of neuropsychological tests in non-English-speaking countries. This study emphasizes that neuropsychological tests commonly used in English-speaking regions and their respective norms should not be directly applied to non-English-speaking individuals.

Acknowledgments

Sponsored by Takeda Development Center Americas, Inc., Deerfield, IL, USA, and Zinfandel Pharmaceuticals, Inc., Durham, NC, USA.

Presented during the Clinical Trials in Alzheimer's Disease (CTAD) Annual Meeting, Barcelona, Spain; November 5-7, 2015.