Variability of Trail Making Test, Symbol Digit Test and Line Trait Test in normal people. A normative study taking into account age-dependent decline and sociobiological variables

Piero Amodio[1], Helmut Wenin[1], Franco Del Piccolo[1], Daniela Mapelli[1,2], Sara Montagnese[1], Andrea Pellegrini[1], Carmine Musto[1], Angelo Gatta[1], and Carlo Umiltà[2]

[1]Department of Clinical and Experimental Medicine, [2]Department of General Psychology, University of Padova, Padova, Italy

ABSTRACT. ***Background and aims:*** *The influence of sociobiological variables and aging on the variability of the Trail Making Tests (TMT), the Symbol Digit Substituting Test (SDT), and the Line Trait Test (LTT) in the general healthy populations are not well known. Even less is known about the reliability at re-testing. This study aimed at determining the reference range of these tests, taking into account sociobiological variables and age, and the re-testing effect.* ***Methods:*** *We studied 300 healthy subjects from 20 to 80 years of age. The sample was derived by the pooling of two samples stratified by age and sex: a randomized sample of 161 subjects collected from the city registers of Padova, and a convenience sample of 139 subjects collected in 20 towns (mainly rural) of Northern Italy. After normalization, data were assayed for the influence of age, education, job, and gender.* ***Results:*** *Age was found to be a significant independent predictor for all the tests, education for all but the LTT, job only for the TMT-B and a geometrical version of the same test (TMT-G) which was proved to be highly correlated with the TMT-B (r=0.80, p<0.01). Job and the interaction age × education level influenced the difference TMT-B minus TMT-A. From the predicting equations, the normative data and the formulas to obtain Z scores for each test were derived. Reliability was lowest for LTT errors (CV=67%), highest for the SDT (13%), whereas the TMT obtained intermediate values (22-33%, depending on the test).* ***Conclusions:*** *This study provides the most reliable normative data range for the TMT, SDT and LTT to date because it considers important demographic variables such as age, education and job.*
(Aging Clin Exp Res 14: 117-131, 2002)

INTRODUCTION

The Trail Making Test (TMT), comprised of two tasks (TMT-A and TMT-B), was developed by Reitan (1) on the basis of the Taylor number series in order to detect organic brain damage (1, 2). Zeegen (3) applied the TMT in evaluating patients who underwent port-caval shunt, and Conn (4) used the TMT-A (renamed Number Connection Test, NCT) as a parameter to quantify hepatic encephalopathy (5). Indeed, the TMT has a wide applicability; it has been used to test early stages of cognitive deterioration (6), dementia (7), cognitive dysfunction in HIV positive patients (8), and in head trauma (9). Nonetheless, demographic factors influence its variability, and their weight in assessing reference ranges is disputed. Parsons et al. (10) found many false positives in normal controls applying Reitan's criteria of normality. Age and education were found to influence the TMT in some studies (10), in others intelligence seemed to be the only significant predictor (11). More recently, Weissenborn et al. (12), Giovagnoli et al. (13) and Amodio et al. (14) suggested a parametric approach based on multivariate regression to define the reference ranges of the TMT. In the study of Weissenborn et al. (12) age, education, and job appeared significant predictors of

Key words: Line Trait Test, healthy people, psychometric performance, reference values, Symbol Digit Test, Trail Making Test.
Correspondence: P. Amodio, M.D., Clinica Medica 5, Via Giustiniani 2, 35128 Padova, Italy. E-mail: piero.amodio@unipd.it
Received August 2, 2001; accepted in revised form December 20, 2001.

TMT variability in the normal population, whereas Giovagnoli et al. (13) considered only age and education.

The Symbol Digit Test (SDT) (15), and similar symbol-substitution tasks like the Digit Symbol Test (16), have also been used to assess the extent of organic brain damage, and to detect and quantify mild hepatic encephalopathy for many years (17-19). A performance decline on the SDT is a well-documented correlate of aging (20). Other recognized variability factors are education (21) and, possibly, gender (22).

More recently, the Line Trait Test (LTT) was introduced by Hamster (23), mainly to study hepatic encephalopathy, and to discriminate alcoholic from liver-related encephalopathy (24). Only age was considered as a variable influencing LTT performance in the normalization studies conducted in Germany (25, 26).

As a whole, despite the popularity of these tests, the factors affecting their variability in the general healthy population (in particular, the Italian population) are not well known, and therefore even less is known about the weight that such factors might have in the assessment of reference values.

The aim of the present study was to determine the major demographic factors predicting the variability of the TMT, SDT, and LTT in normal subjects. We also assessed the applicability of a geometric variant of the TMT-B in which the alphabet sequence is substituted by a geometrical one; this variation may overcome problems arising from the use of the TMT-B alphabetic sequence in modern multiethnic societies in which people with poor knowledge of the Roman alphabet is common (18, 27).

METHODS

Participants

Two samples were considered:

1. A randomized sample collected from the electoral register of the city of Padova by random numbers was stratified according to sex and age in order to have participants for each of the following age classes: 18-30 years, 31-40 years, 41-50 years, 51-60 years, 61-70 years, and 71-80 years. For each age group, 60 subjects (30 males and 30 females) were selected; the first 30 subjects (balanced by gender) constituted the interviewed group, while the second 30 subjects served as a reserve for substituting non-cooperative or excluded subjects.

2. A convenience sample from 20 towns and rural villages of North Italy was collected among the acquaintances of the students of the University of Padova and referred to our team.

The subjects of both groups underwent a structured interview reporting demographic data (birth date, sex, education, job, marital status) and medical history. Job was classified as "blue collar" and "white collar". "Blue collar" workers consisted of craftsmen, farmers, housewives, nurses and hospital technical staff. Their daily work was predominantly manual. "White collar" workers consisted of clerks, students, technical assistants, tradesman, secretaries and university graduates. Additionally, as regards the Italian education system, 4 groups were defined according to years of formal education: at least 5 years (grade of education: 1); at least 8 years (grade of education: 2); at least 13 years (grade of education: 3); and qualification for a university degree (17 years, grade of education: 4). However, all subjects were required to have a fair knowledge of the numerical and Italian alphabetical sequence.

Exclusion criteria were: subjects not found at home after five attempts; refusal to participate in the study; less than 5 years of education (elementary school in Italy); alcohol consumption >70 g/day for males or 40 g/day for females; severe hypertension (duration greater than 5 years and requiring two or more drugs); history of coronary heart disease or cerebrovascular disease, insulin-treated diabetes, severe renal, liver or pulmonary disease; psychiatric history or consumption of psychotropic drugs; history of any kind of cerebral disease. Following their informed consent, each participating subject underwent the 5 paper and pencil psychometric tests in random sequence.

Psychometric tests

The five tests were: 1) the TMT-A, 2) the TMT-B, 3) a geometric version of the TMT-B, that we named TMT-G, 4) the SDT, and 5) the LTT.

1) The TMT-A (1, 2) consisted of 25 circles with a diameter of 2 cm (thickness 0.2 mm) numbered from 1 to 25, written with the font Arial 24 (height 5 mm, thickness 0.7 mm) (Fig. 1). An initial demonstration was performed to familiarize the subjects with the test, then a variation of equal difficulty was used. The subject's task was to connect the circles in sequential order as rapidly as possible.

2) The TMT-B (1, 2) consisted of 25 circles with a diameter of 2 cm containing numbers from 1 to 13 and 12 letters from A to N, all written using the font Arial 24 (Fig. 1). The letters "M" and "N" substituted for the letters "J" and "K" of the original form because the Italian alphabet does not con-

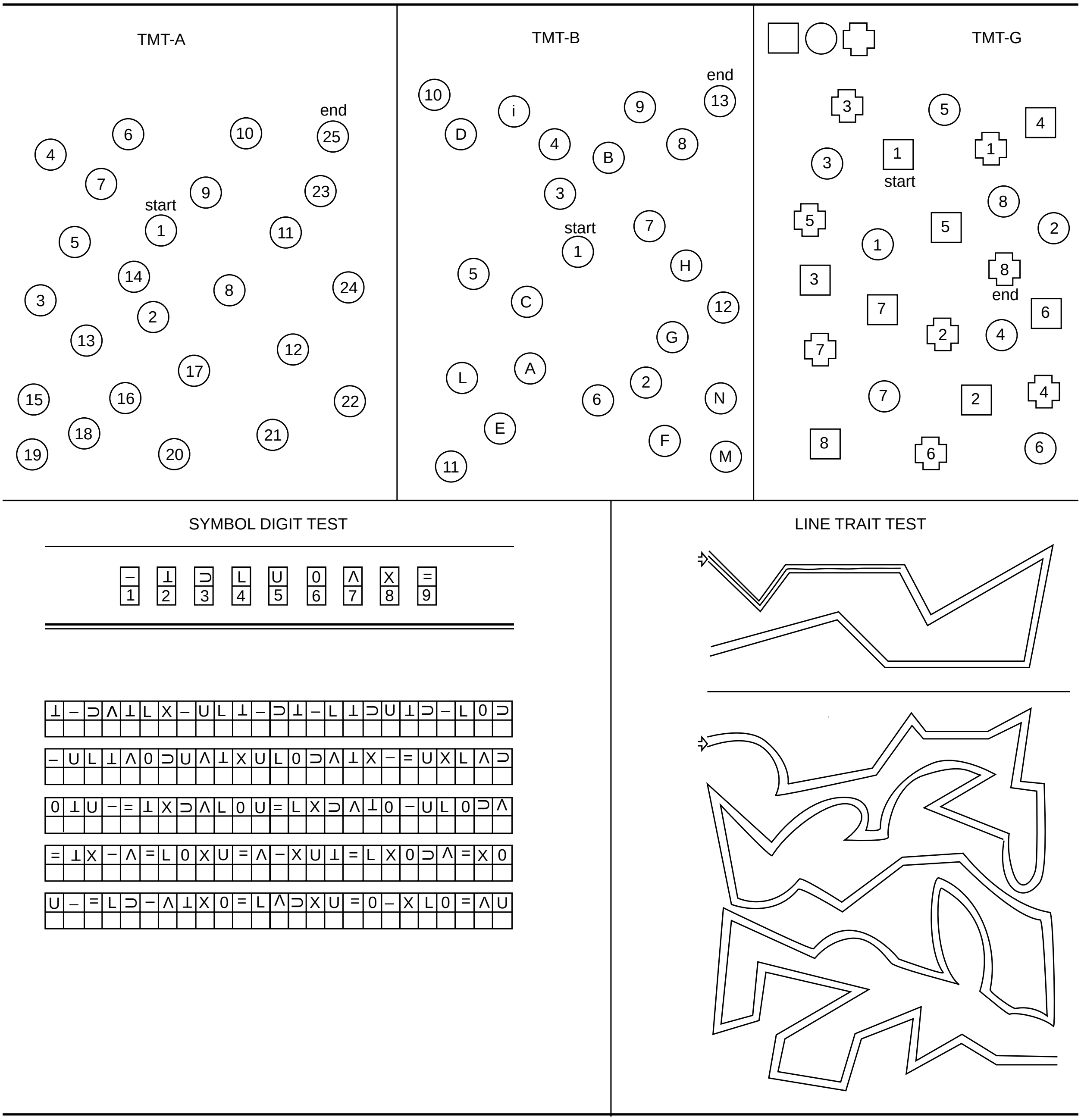

Figure 1 - *The tests used in the study. On the top: the TMT-A, the TMT-B, and the TMT-G; on the bottom: the SDT and the LTT.*

tain these letters. The task was to connect the circles in sequential order alternating between numbers and letters as rapidly as possible.

3) The TMT-G consisted of 8 circles (diameter 1.75 cm, thickness 0.2 mm), 8 squares (side 1.85 cm), and 8 crosses (height 1.95 cm), each containing a number from 1 to 8 written in font Arial 24. The figures and numbers were placed on an A4 sheet using random numbers (Fig. 1). The task was to connect the circle, the square, and the cross containing the number 1, then the circle, the square and the cross containing the number 2, and so on

up to number 8 as rapidly as possible. In this way, the subject has to pay attention both to the sequence of the geometric figures and to that of the numbers. The subject was aided because the correct sequence of geometric figures was reported on the top of the form.

In the TMT-A, TMT-B and TMT-G (TMTs), when the subject makes an error, he/she is required to correct it and continue in the proper sequence. Errors therefore add to the overall time required to complete the task.

Moreover, the differences between TMT-B minus TMT-A (TMT_{B-A}) and the difference between TMT-G minus TMT-A (TMT_{G-A}) were examined, since they depend on shifting ability, reflecting attention capacity (28).

4) The SDT (15, 29) consisted of 5 rows each containing 25 cases for a total of 125 cases (squares of 14 × 7 mm); each case contained one of nine symbols in the top half of the square and was empty in the bottom half. On the top of the form, there was the key to the test: 9 cases containing the nine symbols on the upper half and the corresponding digits on the bottom half (Fig. 1). The character used was Arial 20. The task was to fill as many as possible of the empty half-squares with the digit corresponding to each symbol in 90 seconds. The operator filled in the first three cases, the subsequent seven served as a practice. Time was measured from the tenth case onward.

5) The LTT (25) consisted of a 4 mm-wide contorted track drawn by two parallel lines (thickness 0.4 mm) (Fig. 1). The task was to draw quickly a line inside the track and avoid touching the borders of the track itself. Both the errors and the time to complete the track were considered. The errors were the number of times that the drawn line touched or crossed the limits of the track, and were counted by a transparent template that allowed a different weight of the errors depending on their amplitude. A variant was introduced with respect to the original LTT evaluation: a variable, time × errors, was used based on the observation that the time and errors in performing the test were linked to each other, i.e., the faster the performance, the greater the number of errors [LTT errors $=135 - 25 \times \mathrm{Ln}$ (LTT time); $r=0.32$, $p<0.001$].

Test reliability

To assess the reliability of the tests, 24 subjects in the convenience sample were re-tested a day after the first trial.

Statistics

The raw data for each test were checked by the Kolmogorov-Smirnov test for normality using Lilliefors probabilities. Since the data of no test fitted the Gaussian distribution, after visual inspection different transform functions were applied to normalize the data for each test (see note on bottom of Table 2).

Test results of the random and the convenience samples were compared by ANCOVA adjusting for age and education.

The education levels effective in explaining the variability of each test were assessed by ANCO-

Table 1 - *Demographic characteristics of the subjects enrolled in the study.*

Age (years)	No. of subjects	Random sample (N=161)					Convenience sample (N=139)				
		Males (%)	Education level * (%)				Males (%)	Education level (%)			
			1	2	3	4		1	2	3	4
18-30	30	50	0	12	66	22	53	0	26	58	16
31-40	47	52	3	21	45	31	44	0	40	40	20
41-50	55	37	4	15	22	59	47	8	36	47	9
51-60	63	48	19	7	26	48	50	25	39	22	14
61-70	54	48	24	10	28	38	55	50	17	33	0
71-80	51	59	24	18	35	23	46	46	15	31	8
Totals	300	48	11	14	38	37	49	18	31	39	12

* Education levels: 1= 5 years, 2= 8 years, 3= 13 years, 4= ≥17 years.

Table 2 - *Comparison of normalized psychometric tests according to education levels.*

	5 years	8 years	13 years	≥17 years	p∞
TMT-A	0.26±0.02$^{\$,\#,\S}$	0.29±0.03*	0.29±0.03*	0.30±0.03*	0.002
TMT-B	0.21±0.02$^{\$,\#,\S}$	0.23±0.02$^{*,\#,\S}$	0.24±0.02$^{*,\$}$	0.25±0.02$^{*,\$}$	0.003
TMT-G	0.20±0.02$^{\$,\#,\S}$	0.23±0.02*	0.24±0.02*	0.24±0.02*	0.001
TMT_{B-A}	4.24±0.66$^{\$,\#,\S}$	3.72±0.53$^{*,\#,\S}$	3.41±0.54$^{*,\$}$	3.37±0.62$^{*,\$}$	0.034
TMT_{G-A}	4.51±0.56$^{\$,\#,\S}$	3.82±0.59*	3.65±0.54*	3.65±0.64*	0.001
SDT	4.87±0.06$^{\$,\#,\S}$	4.75±0.10$^{*,\#,\S}$	4.69±0.10$^{*,\$}$	4.69±0.10$^{*,\$}$	0.001
LTT-t	4.45±0.36$^{\#}$	4.27±0.30	4.27±0.34	4.37±0.32	0.33
LTT-er	3.52±0.80$^{\#,\S}$	3.23±0.70	3.05±0.80	2.97±0.87	0.16
LTT-wt	5.69±0.31$^{\$,\#,\S}$	5.41±0.29*	5.34±0.35	5.42±0.31	0.58

*∞ p-values adjusted for age and job (ANCOVA). p<0.05 with respect to education of 5 years *, 8 years $, 13 years #, ≥17 years § (Scheffé test for multiple comparison was performed when ANCOVA was significant). t: time; er: errors; wt: weighted time; e: number of Euler, i.e., 2.71828.*

Note: Transform functions applied for normalizing the distributions of crude psychometric test results: TMT-A: 1/Ln (TMT-A). TMT-B: 1/Ln (TMT-B). TMT-G: 1/Ln (TMT-G). TMT-B minus TMT-A: Ln (TMT_{B-A}). TMT-G minus TMT-A: Ln (TMT_{G-A}). SDT: Ln (160-SDT). LTT-t: Ln (LTTt). LTT-er: Ln (e + LTT-er). LTT-wt: Ln LTT-wt

VA, adjusting for age and job (blue *vs* white collars), and *post-hoc* assessment by the Scheffé test.

The role of age, education levels, job and gender to explain the variability of each test was assessed by multiple regression. Moreover, a variable indicating the group of origin (convenience or randomized) was used to verify that the pooling of the two samples was correct.

The reference ranges of the tests were calculated from the predictive equations, adjusting the standard errors by functions that allow a better fit to the empirical findings.

Reliability was measured by the coefficient of repeatability according to Bland and Altman (30), i.e., by the standard deviations of the repeated measuring, and by a coefficient of variability given by the ratio of the coefficient of repeatability to the mean value of the test. This technique provides clear information, since a repeated measure has 95% probability of falling in the interval twice the coefficient of repeatability with respect to the first measure (30). In addition, Pearson's coefficient of correlation between repeated measures was calculated.

The package "Statistica 5.5" (StatSoft Inc., Tulsa, Oklahoma, USA) was used for the statistical analysis.

RESULTS

Sample

Of the 360 subjects randomized from the electoral register of the city of Padua, 161 were enrolled in the study. One hundred ninety-nine subjects were excluded: 1 for education less than 5 years, 37 for severe diseases or psychotropic drug consumption, 4 for high alcohol consumption, 42 because they had changed their residence, 100 denied their consent, and 6 because their age group had been already completed.

The convenience sample comprised 139 subjects fulfilling the same criteria adopted for the randomized sample. To collect 139 subjects, 144 subjects were considered: 5 were excluded because an accurate interview disclosed alcohol abuse or diseases considered in the exclusion list.

The demographic characteristics of the subjects enrolled in the study are reported in Table 1.

Variability of psychometric tests and the reference values

The raw data did not show statistically significant differences between the randomized and the convenience samples, except for the time of performance of the LTT which was statistically, albeit negligibly, higher in the random sample than in the convenience sample (84±29 *vs* 74±29 sec, $p<0.001$) where people seemed to give higher value to accuracy (27±28 *vs* 31±21 errors, $p=0.23$). When the time of performance was weighted for accuracy, there was no difference in the two samples (230±81 *vs* 223±75 sec × errors, $p=0.43$). Therefore the data of the randomized and convenient samples were pooled together.

The distribution of the data for each test showed non-normality, therefore each test was normalized according to an appropriate transform function (see note on bottom of Table 2).

When adjusted for age and job, education level was found to influence the TMTs, TMT_{B-A}, TMT_{G-A} and SDT, but not the LTT (Table 2). The performances of the TMT-A, TMT-G, and TMT_{G-A} were significantly lower in the subjects who had only 5 years of education (elementary school), whereas those of the TMT-B, TMT_{B-A} and SDT were lower in the subjects who had 5 years of education, intermediate in the subjects who had 8 years of education, and higher in the subjects who had 13 years of education (high school) or more (University degree) (Table 2).

Variables to indicate sampling (convenience or randomized sampling), education levels, job, age, and gender, and their interactions were used as predictors in multiple regression models to assess the combined influence that they may have in explaining the variability of psychometric tests. Of the predictors considered, only age was found to explain significantly the variability of all the tests. In addition, education influenced the TMTs ($p<0.01$) and SDT ($p<0.001$), while job influenced only the TMT-B ($p<0.001$) and TMT-G ($p<0.01$). Gender never entered as an independent predictor of any test. An interaction between age and education in the TMT_{B-A} was found ($p<0.001$).

In order to define the cut-off values, we calculated the regression equations and the standard deviations (SD) for each test according to the predictors found to be significant, using differing equations for the various education levels. However, the use of the mean +2 SD of the expected values to define the upper limit of normality exaggerates the widening effect for the upper end of the model (i.e., in older people) due to the exponential models used to transform the raw data. Therefore, to fit the empirical data and define the limits of normal values in healthy people, we applied equations progressively reducing the exponential increase of the SD due to age progression (Tables 3 and 4, and Appendix). Visual inspection was applied to verify these concepts (Appendix). Applying this principle, the adjusted value of a psychometric test can be expressed as the standardized residual from the actual value and the value expected by age, and possibly other confounders (education and job), according to the equations shown in the Appendix.

Table 3 - *The limits of the reference range for the TMTs, LTT, and SDT.*

Age	TMT-A (sec)		TMT-B (sec)				TMT-G (sec)			SDT (items/90 sec)			LTT-t (sec)	LTT-er	LTT-wt
years	5 y.*	≥8 y.	5 y.	8-13 y.	≥13 y. Blue collar	≥13 y. White collar	5 y	≥8 y. Blue collar	≥8 y White collar	5 y.	8-13 y.	≥13 y.			
20	46	39	100	93	105	82	136	100	88	34	40	44	115	61	261
25	49	42	111	101	113	88	148	109	96	32	37	42	120	67	279
30	53	44	122	109	122	94	163	119	104	30	35	40	124	73	298
35	57	47	136	119	133	101	180	130	114	28	32	38	128	81	319
40	62	51	151	130	144	109	199	143	124	25	29	36	133	88	342
45	67	55	169	142	156	117	220	158	136	23	27	33	138	97	365
50	72	59	189	155	170	126	244	174	149	21	24	31	143	105	391
55	78	63	211	169	184	135	271	191	163	19	21	29	148	114	418
60	83	67	236	184	199	145	299	210	178	17	19	27	153	123	447
65	88	72	263	199	213	154	329	230	194	15	16	25	159	132	479
70	92	76	291	214	227	163	356	249	208	13	14	23	165	141	512
75	95	79	318	227	239	171	379	265	221	11	12	21	170	150	548
80	95	81	343	238	248	176	394	276	230	9	9	19	177	157	586

*y.= years of education.

Table 4 - *The limits of the reference range for the TMT_{B-A} and the TMT_{G-A}.*

Age	TMT_{G-A} (sec)		TMT_{B-A} (sec)				
years	5 y.*	≥8 y.	5 y.	8-13 y. Blue collar	8-13 y. White collar	≥13 y. Blue collar	≥13 y.White collar
20	104	72	70	63	55	74	55
25	115	78	79	69	58	79	58
30	126	86	90	77	61	84	61
35	139	93	102	85	67	88	65
40	152	102	115	94	74	93	68
45	167	111	129	104	82	97	71
50	182	121	145	115	90	102	75
55	199	131	163	125	99	107	78
60	217	141	182	136	107	111	81
65	236	151	203	147	115	115	85
70	257	161	226	156	123	120	88
75	279	169	250	164	129	124	91
80	302	177	277	169	133	128	94

*y.= years of education.

Relationship across psychometric tests

A general view of the relationships among these psychometric tests is given in Table 5. The TMT-B and TMT-G were found to be closely linked (r=0.8, $p<0.0001$), confirming that it was reasonable to assume that they reflect, at least in part, analogous cognitive functions. Examination of the LTT disclosed that neither LTT performing time (LTTt) or LTT errors (LTT-er) alone appeared to be linked with the other psychometric variables, whereas the LTT time weighted (LTT-wt) by the errors appeared more consistent with the other psychometric tests.

Table 5 - *Matrix of correlations (Parson's r) across the TMTs, LTT and SDT.*

	TMT-A	TMT-B	TMT-G	TMT_{B-A}	TMT_{G-A}	SDT	LTT-t	LTT-er
TMT-B	0.74 *p*=0.000							
TMT-G	0.71 *p*=0.000	0.80 *p*=0.000						
TMT_{B-A}	0.45 *p*=0.000	0.85 *p*=0.000	0.65 *p*=0.000					
TMT_{G-A}	0.49 *p*=0.000	0.66 *p*=0.000	0.85 *p*=0.000	0.68 *p*=0.000				
SDT	-0.68 *p*=0.000	-0.75 *p*=0.000	-0.75 *p*=0.000	-0.67 *p*=0.000	-0.67 *p*=0.000			
LTT-t	0.38 *p*=0.000	0.37 *p*=0.000	0.31 *p*=0.000	0.32 *p*=0.000	0.25 *p*=0.000	-0.33 *p*=0.000		
LTT-er	0.21 *p*=0.002	0.21 *p*=0.001	0.26 *p*=0.000	0.17 *p*=0.003	0.22 *p*=0.000	-0.35 *p*=0.000	-0.42 *p*=0.000	
LTT-wt	0.53 *p*=0.000	0.54 *p*=0.000	0.51 *p*=0.000	0.46 *p*=0.000	0.44 *p*=0.000	-0.62 *p*=0.000	0.60 *p*=0.000	0.43 *p*=0.000

t: time; er: errors; wt: performing time weighted by the errors.

Table 6 - *The coefficients of correlation, repeatability, and variations between repeated measures of the TMTs, SDT and LTT.*

Test	Coefficient of correlation	Coefficient of repeatability	Coefficient of variation
TMT-A	r=0.87 *p*<0.001	8 sec	24%
TMT-B	r=0.86 *p*<0.001	22 sec	33%
TMT-G	r=0.89 *p*<0.001	16 sec	22%
TMT_{B-A}	r=0.82 *p*<0.001	18 sec	54%
TMT_{G-A}	r=0.74 *p*<0.001	17 sec	40%
SDT	r=0.93 *p*<0.001	7 items	13%
LTT-er	r=0.67 *p*<0.001	18 errors	67%
LTT-t	r=0.65 *p*=0.001	19 sec	24%
LTT-wt	r=0.86 *p*<0.001	40 sec	18%

Test reliability

Reliability of the tests was evaluated by the coefficients of correlation, the coefficient of repeatability (CR) and the coefficient of variability (CV). Reliability was higher for the SDT and LTT-wt; medium for TMT-A, TMT-G, and LTT-t; lower for the TMT-B, TMT_{B-A}, TMT_{G-A}, and LTT-er (Table 6).

DISCUSSION

The agreement between the psychometric tests in the randomized sample collected in the city and the convenience sample collected in villages suggests the consistency and validity of the samples. Certainly, a much more accurate sampling technique would use the randomization for the entire sample, but it is worth noting that exact sampling techniques have not been applied in any of the previous studies on the TMTs, the LTT, or the SDT. All were performed on convenience samples, and some even with paid volunteers, a technique that may easily cause selection bias. Therefore, even if not unquestionable, our data may represent an improvement in sampling collection and provide a better representation of the general healthy population across a wider age range than that considered in the majority of the previous studies.

In disagreement with Boll and Reitan (11), we confirmed the studies (12-14, 21, 31-34) showing the influence of age on the TMTs. The importance of aging in psychometric performance was also confirmed for the SDT and LTT, in agreement with previous observations (8, 15, 25, 26, 35). Moreover, the direct correlations of age with TMT-B minus TMT-A, and TMT-G minus TMT-A showed that aging more relevantly impairs the tasks needing a higher attention and/or working memory capacity. This finding agrees with the observation of Cabeza et al. (36), suggesting that aging selectively decreases the metabolic activity of pre-frontal areas subserving these cognitive functions.

Of the other variability factors, education was confirmed to be an important confounder of the TMTs and SDT, in agreement with previous studies (12, 13, 21, 37). In addition, we observed that people with 8 or 13 years of education (in relation to the test considered) not only performed the tests definitely better than people with a lower education level, but also reached a plateau of performance. This finding means that the years of education did not have a linear effect on psychometric performance. Moreover, that no interaction was found between education level and age (as reinforced by the parallelism of the curves of psychometric decline as a function of age in subjects with different education levels) in all tests (but in the difference TMT-B minus TMT-A) is in line with current knowledge that education generally does not influence the rate of age-related cognitive decline (38-40). However, an important exception was given by the difference TMT-B minus TMT-A which evidenced a lower age-related rate of cognitive decline in high-educated subjects. This finding might be explained by a training effect on alphabet sequence in high-educated subjects. In any case, if confirmed, this finding would represent an important exception to the theory that age-related cognitive decline is a mere biological process that cannot be influenced by sociobiographical variables (39).

The role of job as an explanatory factor of TMTs, SDT, or LTT variability has not been considered previously, except for the study of Weissenborn et al. (12). We found that job has a role as performance predictor for the TMT-B and TMT-G, which are the two similar and highly correlated tests with higher sustained attention loads. It is conceivable that the kind of job may have a training effect on sustained attention and/or working memory. Nevertheless, job, compared to age and education, appeared only as a minor, possibly negligible, predictor of the TMT-B and TMT-G, as found by Weissenborn et al. (12). The lack of any influence of gender in the performance of the TMTs, SDT, and LTT was in line with the majority of the studies (20, 35, 41), except for a few references suggesting better performance of the SDT (21, 22) and TMTs (33) in women. However, findings based on these last studies are questionable, because they were derived from convenience samples (some on paid volunteers). In addition, the technique used to adjust for confounders that may account for gender difference was not always adequate.

Due to the transforms used to normalize the tests, the effect of the predictors on the tests was exponential. Based on visual inspection, this modelling was quite satisfactory for the curve that fitted the mean values of the age-related psychometric decay. However, in this model the extremes of a reference range defined by the mean +2 SD increased exaggeratedly with aging, so that they appeared to be clearly overestimated in older people. On closer inspection, this overestimation is also present in the study of Weissenborn et al. (12) who used an exponential model. For this reason, a better fit may be given by the introduction of an age-dependent reducing coefficient that decreases the spread of the upper limits of the reference ranges in old people. Nevertheless, such an approach to the problem of describing the variability of a parameter in a population is rather atypical, because it gives information derived from a definite function, as in parametric models, as well as the empirical observation of data dispersion, as in non-parametric approaches. In this way, the expression of the results by Z-values, though adjusted for age, is still possible. Indeed, the expression of results as Z-values (SD) from those expected for age, education, and possibly job may simplify their interpretation and comparison, avoiding the need to subdivide results into subgroups (14). In such a way, the rather crude distinction between tests performed within or without a reference range is overcome, because the use of a Z-score immediately provides information on the spread from the expected values in the reference population.

A strict comparison of the normative data of the present study with those of previous studies on the TMTs, SDT and LTT is not possible, because details concerning the forms of psychometric tests and sample collection are generally lacking, as well as proper handling of confounders. Moreover, the range of age that we considered is wider than that of previous research. However, a rough comparison showed that our normative data fell in line with those reported in previous studies, at least in the age groups where the comparison was possible. More in detail, our data showed a slightly quicker performance of the TMTs compared to the data of Giovagnoli et al. (13). LTT performance was slower but more accurate than the findings of Ennen (26), but faster and less accurate compared with the findings of Hamster et al. (23). Moreover, data on the LTT-wt have never been reported before, notwithstanding the fact that its link with the other psychometric tests suggests that it provides more relevant psychometric information than the time or the errors in performing the LLT by themselves.

As regards the reliability of the tests, to our knowledge no study adequately considered this point. Even the study of Giovagnoli et al. (13), which considers the repeatability of the TMT, only applied a statistical tool (i.e., the correlation) that merely indicated the non-random relationship between the first and second psychometric session. Our study confirmed that the tests were well correlated on re-testing; in addition, the use of the coefficient of repeatability provided a reliable measure of variability that allowed assessment of the significance of a difference between repeated measures (30). In this way, it was evidenced that small variations in repeated measures of single subjects should be interpreted cautiously, because intra-subject variability is not negligible.

In conclusion, our study provides the most reliable normative data range to date for the TMTs, SDT and LTT, which takes into account important demographic variables such as age, education and job.

APPENDIX

Calculation of the adjusted Z-score considering the main confounders for each psychometric test.

For each test the Z-scores adjusted for the main confounders were calculated by the difference between the performance expected on the basis of the predictive regression and the observed performance, divided by the standard errors of the predictive regressions. Different regressions were applied for each education level. The standard errors were adjusted by functions reducing the age-dependent exponential increase of the errors, when needed. The goodness of the adjusting functions was assessed by visual inspection (Appendix Figures 1-7).

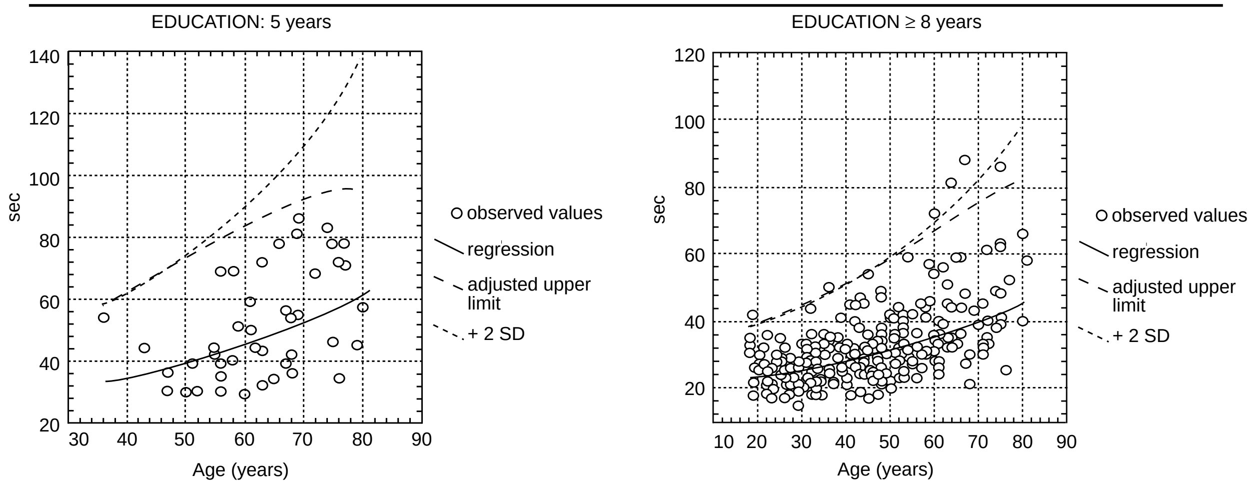

Figure 1 - *Relationships between TMT-A, age, and education. The circles show the observed values, the continuous line shows the mean square fitting, the line with shorter dashes represents the expected upper limit, without adjustment, the line with longer dashes represents the adjusted upper limit.*

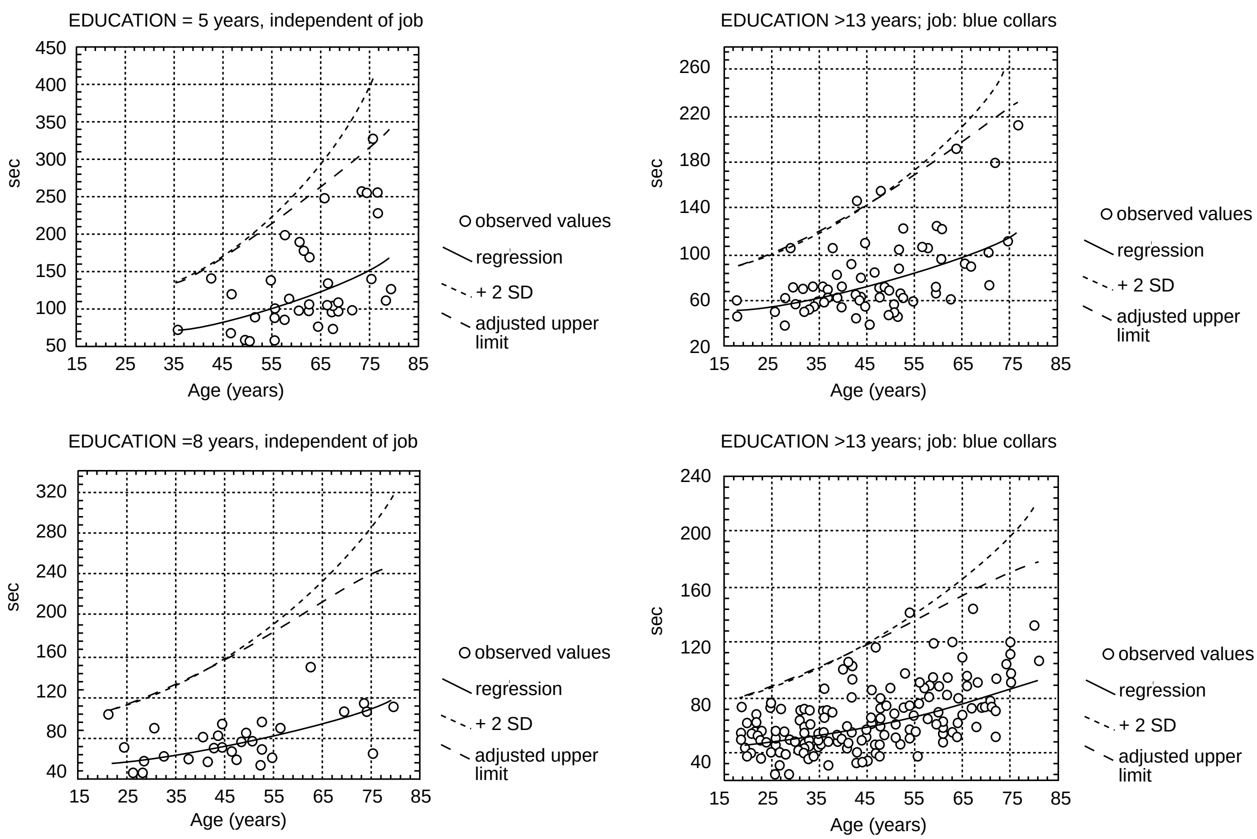

Figure 2 - *Relationships between TMT-B, age, education, and job. The circles show the observed values, the continuous line shows the mean square fitting, the line with shorter dashes represents the expected upper limit, without adjustment, the line with longer dashes represents the adjusted upper limit.*

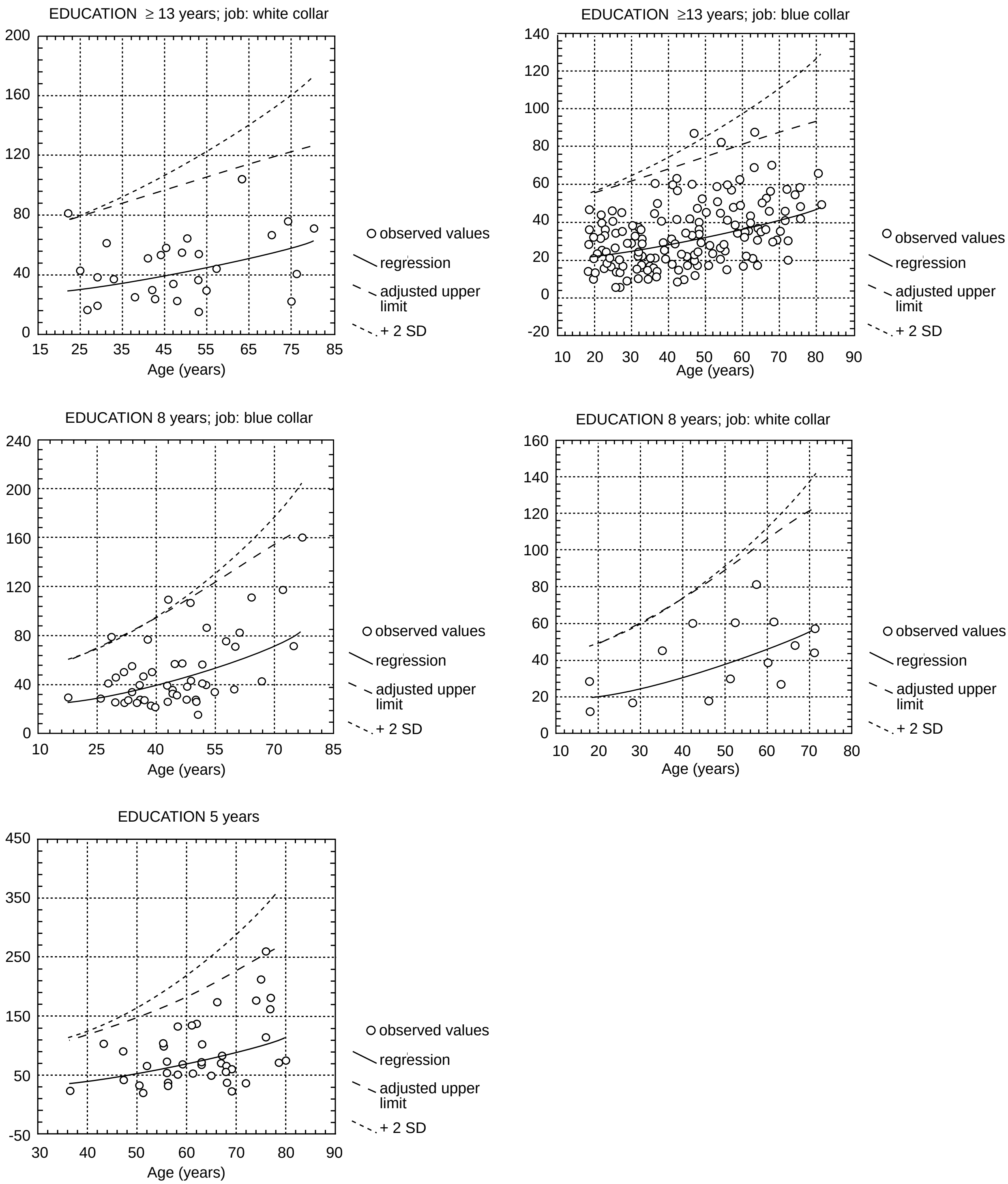

Figure 3 - *Relationships between the difference TMT-B minus TMT-A, age, education, and job. The circles show the observed values, the continuous line shows the mean square fitting, the line with shorter dashes represents the expected upper limit, without adjustment, the line with longer dashes represents the adjusted upper limit.*

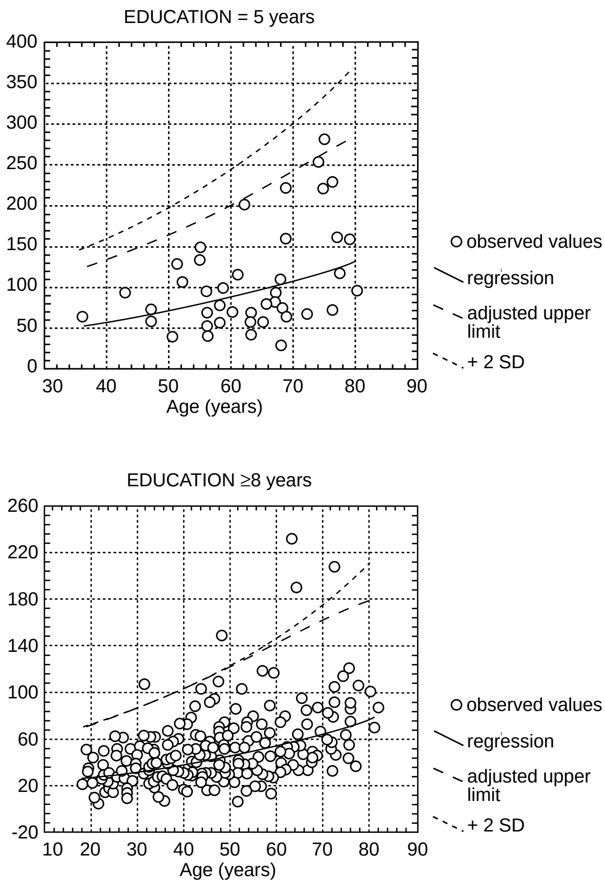

Figure 4 - *Relationships between the difference TMT-G minus TMT-A, age, education. The circles show the observed values, the continuous line shows the mean square fitting, the line with shorter dashes represents the expected upper limit, without adjustment, the line with longer dashes represents the adjusted upper limit.*

TMT-A

Education = 5 years

$Z=[(0.3212-0.00098 \times \text{age})-1/\text{Log(TMT-A)}]/(0.02-\text{age}^5/120 \times 80^5)$;

Education ≥8 years

$Z=[(0.3356-0.000914 \times \text{age})-1/\text{Log(TMT-A)}]/(0.0223-\text{age}^7/200 \times 80^7)$

TMT-B

Education = 5 years

$Z=[(0.267-0.0009 \times \text{age})-1/\text{Log (TMT-B)}]/(0.016-\text{age}^5/3 \times 80^6)$

Education ≥8 years

$Z=[(0.268-0.00077 \times \text{age})-1/\text{Log (TMT-B)}]/(0.016-\text{age}^5/3 \times 80^6)$

Education ≥13 years

$Z=[(0.265-0.0007 \times \text{age}+0.012 \times \text{job})-1/\text{Log (TMT-B)}]/(0.018-\text{age}^5/3 \times 80^6)$

TMT-G

Education = 5 years

$Z=[(0.25-0.000745 \times \text{age})-1/\text{Log (TMT-G)}]/(0.0157-\text{age}^6/3 \times 80^7)$

Education ≥8 years

$Z=[(0.266-0.000792 \times \text{age}+0.00598 \times \text{job})-1/\text{Log (TMT-G)}]/(0.0165-\text{age}^6/3 \times 80^7)$

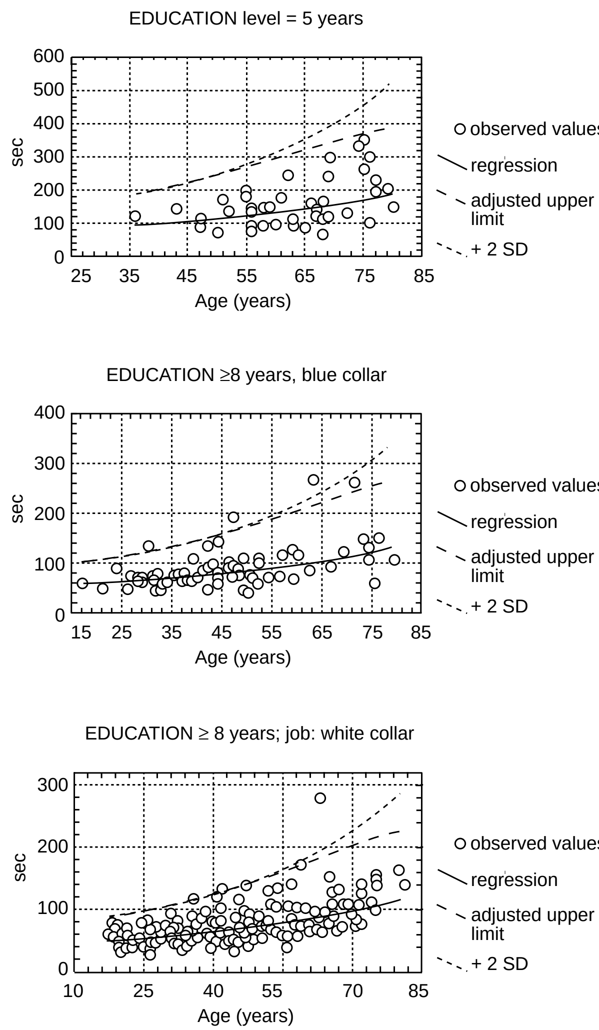

Figure 5 - *Relationships between TMT-G, age, education, and job. The circles show the observed values, the continuous line shows the mean square fitting, the line with shorter dashes represents the expected upper limit, without adjustment, the line with longer dashes represents the adjusted upper limit.*

Difference TMT-B minus TMT-A

Education 5 years

$Z=[\text{Log }(\text{TMT-B} - \text{TMT-A})-(2.5+0.028 \times \text{age})]/(0.6022-\text{age}^2/200^2)$

Education 8 years

$Z=[\text{Log }(\text{TMT-B} - \text{TMT-A})-(2.81+0.021 \times \text{age} -0.24 \times \text{job})]/(0.453-\text{age}^5/120^5)$

where: job = 0 for blue collars
job = 1 for white collars

Education ≥13 years

$Z=[\text{Log }(\text{TMT-B} - \text{TMT-A})-(3.03+0.014 \times \text{age} -0.31 \times \text{job})]/(0.5102-\text{age}^2/200^2)$

where: job = 0 for blue collars
job = 1 for white collars

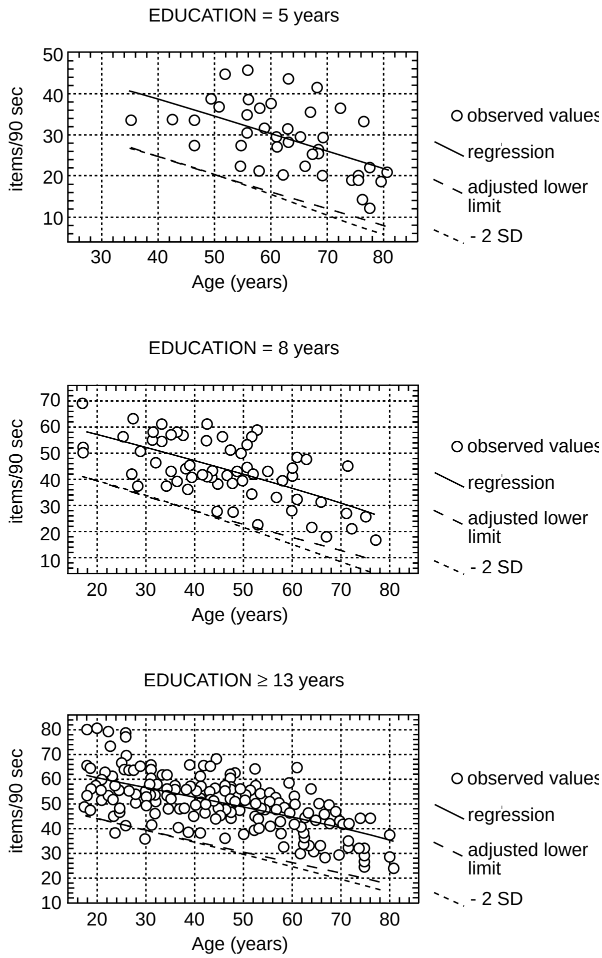

Figure 6 - *Relationships between SDT, age, and education. The circles show the observed values, the continuous line shows the mean square fitting, the line with shorter dashes represents the expected lower limit, without adjustment, the line with longer dashes represents the adjusted lower limit.*

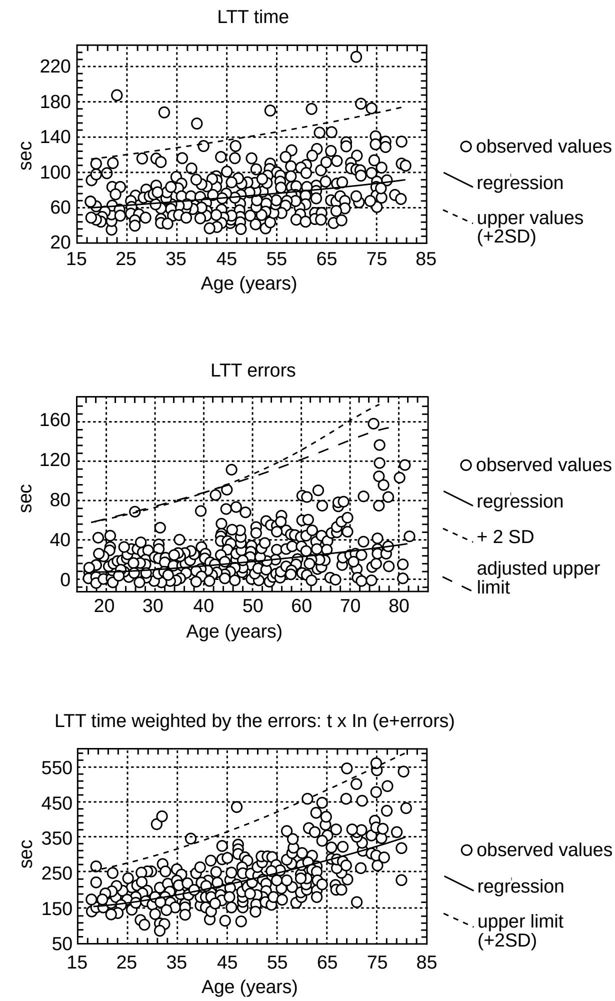

Figure 7 - *Relationships between LTT and age. The circles show the observed values, the continuous line shows the mean square fitting, the line with shorter dashes represents the expected upper limit, without adjustment, the line with longer dashes represents the adjusted upper limit.*

Difference TMT-G minus TMT-A

Education 5 years

$Z=[\text{Log }(\text{TMT-G} - \text{TMT-A})-(3.2+0.021 \times \text{age})]/(0.518-\text{age}^2/250^2)$

Education ≥8 years

$Z=[\text{Log }(\text{TMT-G} - \text{TMT-A})-(2.876+0.018 \times \text{age})]/(0.517-\text{age}^5/130^5)$

SDT

Education = 5 years

$Z=[\text{Log}(-(\text{SDT-160}))-(4.667+0.0032 \times \text{age})]/(0.054-\text{age}^4/90^5)$,

Education 8-13 years

$Z=[\text{Log }(-(\text{SDT-160}))-(4.545+0.0044 \times \text{age})]/(0.078-\text{age}^{3.1}/80^4)$, *Education ≥13 years*

Z=[Log(-(SDT-160))-(4.523+0.00374 × age)]/(0.077-age^3/80^4)

LTT-time

Z=[Log (LTT)-(3.98+0.0071 × age)]/0.313

LTT-errors

Z=[Log(e+LTT-er)-(2.248+0.0188 × age)]/(0.762-age^4/10 × 80^4)

LTT-time weighted by the errors

Z={Log[Log(e+LTT-er) × LTT-t]-(4.776+0.0135 × age)}/0.258

ACKNOWLEDGEMENTS

We thank Anthony Schork, Professor Emeritus at the University of Michigan (USA) and Adjunct Statistician at the University of Padova (Italy) for reviewing the manuscript. This work was supported in part by the Italian Liver Foundation.

REFERENCES

1. Reitan RM. The relation of the Trail Making Test to organic brain damage. J Consult Psychol 1955; 19: 393-4.
2. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958; 8: 271-6.
3. Zeegen R, Drinkwater JE, Dawson AM. Method for measuring cerebral dysfunction in patients with liver disease. BMJ 1970; 2: 633-6.
4. Conn HO. Trailmaking and number-connection tests in the assessment of mental state in portal systemic encephalopathy. Am J Dig Dis 1977; 22: 541-50.
5. Conn HO, Leevy CM, Vlahevic ZR, et al. Comparison of lactulose and neomicin in the treatment of chronic portal-sistemic encephalopathy. Gastroenterology 1977; 72: 573-83.
6. Greenlief CL, Margolis RB, Erker GJ. Application of the trail making test in differentiating neuropsychological impairment of elderly persons. Percept Mot Skills 1985; 61: 1283-9.
7. Rasmusson DX, Zonderman AB, Kawas CH, Resnick SM. Effects of age and dementia on the Trail Making Test. Clin Neuropsychol 1998; 2: 169-78.
8. Selnes OA, Jacobson L, Machado AM, et al. Normative data for a brief neuropsychological screening battery. Multicenter AIDS Cohort Study. Percept Mot Skills 1991; 73: 539-50.
9. Leininger BE, Gramling SE, Farrell AD, Kreutzer JS, Peck EA. Neuropsychological deficits in symptomatic minor head injury after concussion and mild concussion. J Neurol Neurosurg Psychiatry 1990; 53: 293-6.
10. Parsons OA, Maslow HI, Morris F. Trail Making Test performance in relation to certain experimenter, test and subject variables. Percept Mot Skills 1964; 19: 199-206.
11. Boll TJ, Reitan RM. Effect of age on performance of the Trail Making Test. Percept Mot Skills 1973; 36: 691-4.
12. Weissenborn K, Ruckert N, Hecker H, Manns M.P. The number connection tests A and B: interindividual variability and use for the assessment of early hepatic encephalopathy. J Hepatol 1998; 28: 646-53.
13. 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: 305-9.
14. Amodio P, Del Piccolo F, Marchetti P, et al. Clinical features and survival of cirrhotic patients with subclinical cognitive alterations detected by the number connection test and computerized psychometric tests. Hepatology 1999; 29: 1662-7.
15. Lezak MD. Neuropsychological Assessment. New York: Oxford University Press, 1995.
16. Wechsler D. Wechsler Adult Intelligence Scale-Revised. San Antonio, TX: The Psychological Corporated, 1981.
17. Morgan SF, Wheelock J. Comparability of WAIS—R Digit Symbol and the Symbol Digit Modalities Test. Percept Mot Skills 1995; 80: 631-4.
18. Quero JC, Hartmann IJ, Meulstee J, Hop WC, Schalm SW. The diagnosis of subclinical hepatic encephalopathy in patients with cirrhosis using neuropsychological tests and automated electroencephalogram analysis. Hepatology 1996; 24: 556-60.
19. Tarter RE, Hegedus AM, Van Thiel DH, Schade RR, Gavaler JS, Starzl TE. Nonalcoholic cirrhosis associated with neuropsychological dysfunction in the absence of overt evidence of hepatic encephalopathy. Gastroenterology 1984; 86: 1421-7.
20. Gilmore GC, Royer FL, Gruhn JJ. Age differences in symbol-digit substitution task performance. J Clin Psychol 1983; 39: 114-24.
21. Mortensen EL, Gade A. On the relation between demographic variables and neuropsychological test performance. Scand J Psychol 1993; 34: 305-17.
22. Portin R, Saarijarvi S, Joukamaa M, Salokangas RK. Education, gender and cognitive performance in a 62-year-old normal population: results from the Turva Project. Psychol Med 1995; 25: 1295-8.
23. Hamster W. Motor Performance Scale (MPS) Manual. Moedling, Austria: Dr. G. Schufried, Scientific Instruments, 1983.
24. Schomerus H, Hamster W. Neuropsychological aspects of portal-systemic encephalopathy. Metab Brain Dis 1998; 13: 361-77.
25. Hamster W, Kluck M, Schomerus H. PSE-Syndrom-Test -Kurzform- Manual. Hannover: Duphar Pharma, 1985.
26. Ennen J. Diagnosestandardisierung der portosystemischen Encephalopathie mittels psychometrischer Testverfahren. Hannover: Medizinische Hochschule, 1999 (Thesis/Dissertation).
27. Dhiman RK, Saraswat VA, Verma M, Naik SR. Figure connection test: a universal test for assessment of mental state. J Gastroenterol Hepatol 1995; 10: 14-23.
28. Corrigan JD, Hinkeldey NS. Relationships between parts A and B of the Trail Making Test. J Clin Psychol 1987; 43: 402-9.
29. Russell EW. A WAIS factor analysis with brain damaged subjects using criterion measures. J Consult Clin Psychol 1972; 39: 133-9.
30. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307-10.
31. Goul WR, Brown M. Effects of age and intelligence on trail making test performance and validity. Percept Mot Skills 1970; 30: 319-26.
32. Davies ADM. The influence of age on Trail Making Test performance. J Clin Psychol 1968; 24: 96-8.
33. Bornstein RA. Normative data on selected neuropsychological measures from a nonclinical sample. J Clin Psychol 1985; 41: 651-9.

34. Nielsen H, Knudsen L, Daugbjerg O. Normative data for eight neuropsychological tests based on a Danish sample. Scand J Psychol 1989; 30: 37-45.

35. Yeudall LT, Fromm D, Reddon JR, Stefanyk WO. Normative data stratified by age and sex for 12 Neuropsychological Tests. J Clin Psychol 1986; 42: 918-46.

36. Cabeza R, Grady CL, Nyberg L, et al. Age-related differences in neural activity during memory encoding and retrieval: a positron emission tomography study. J Neurosci 1997; 17: 391-400.

37. Amodio P, Quero JC, Del Piccolo F, Gatta A, Schalm SW. Diagnostic tools for the detection of subclinical hepatic encephalopathy: comparison of standard and computerized psychometric tests with spectral-EEG. Metab Brain Dis 1996; 11: 315-27.

38. del Ser T, Hachinski V, Merskey H, Munoz DG. An autopsy-verified study of the effect of education on degenerative dementia. Brain 1999; 122: 2309-19.

39. Park DC. The basic mechanisms accounting for age-related decline in cognitive function. In Park DC, Schwarz N, Eds. Cognitive Aging: A Primer. Philadelphia: Psychology Press, 2001: 3-21.

40. Lindenberger U, Baltes PB. Intellectual functioning in old and very old age: cross-sectional results from the Berlin Aging Study. Psychol Aging 1997; 12: 410-32.

41. Waldmann BW, Dickson AL, Monahan MC, Kazelskis R. The relationship between intellectual ability and adult performance on the Trail Making Test and the Symbol Digit Modalities Test. J Clin Psychol 1992; 48: 360-3.