Norms for Older New Zealanders on the Trail-Making Test

Richard J Siegert
Victoria University of Wellington

Carolyn M Cavana
Gilmour, McGregor and Associates Ltd

Despite frequent calls for the development of appropriate norms for older people for many widely used neuropsychological tests, there remains a paucity of such data. In the present study we collected norms for the Trail-Making Test based upon the performance of 127 New Zealanders aged 60 years or older. The Trail-Making Test is a popular test with clinicians for the assessment of attentional difficulties, combining ease of administration, economy of time and money, and palatability to patients. In addition to providing norms for 60 to 85 year olds on this test, regression equations that allowed us to give expected scores on the Trail-Making Test taking intellectual ability into account, were developed. The results underline the importance of using appropriate age-based norms in judging performance and making clinical decisions. The added importance of considering intelligence is also demonstrated.

Improved living standards and developments in medical technology have resulted in a profound alteration of the age structure of the populations of most developed countries in the late twentieth century. Perhaps the most notable change has been an increase in the proportion of the population over retirement age (Richmond, Baskett, Bonita, & Melding, 1995). Indeed, it is the so-called 'old elderly' (those aged 75 years and over) who comprise the fastest growing sector of the population. Hart and Semple (1990) comment in this regard that in the UK retired pensioners currently outnumber children of school age. While the majority of older people continue to lead independent and active lives, a rapid increase in the prevalence of age-related cognitive disorders has been noted for several years (Lock, 1978). Clinical psychology and, in particular, neuropsychology, has an essential role to play in the provision of health care and support to the elderly and their relatives coping with the impact of these neurological illnesses. A thorough cognitive assessment can be useful in differential diagnosis, assessing the progress of a degenerative illness, and in planning and assessing rehabilitation programmes (Cipolotti & Warrington, 1995). Perhaps the most exciting new role for neuropsychological assessment will be in the assessment of the effectiveness of new pharmocological interventions for degenerative diseases (Hart & Semple, 1990).

In the clinical neuropsychological assessment of older adults' cognitive functioning, there are two important sources of information. The first is information from the person's spouse or relatives. The second is a comprehensive assessment including an interview and appropriate psychometric testing. However, interpretation of cognitive function presupposes a notion of normative age-related functioning. A common problem confronting the clinician is how to differentiate acquired impairment from changes related to normal aging. This is especially true in the early stages of disorders such as Dementia of the Alzheimer's Type (DAT) where the onset is typically insidious and symptoms such as forgetfulness are common in the non-affected elderly. Several researchers have commented upon the need for age-appropriate normative data (Bolla Wilson & Bleecker, 1986; Bornstein, 1985; Prigatano 1978).

Lezak (1987) considered the age norm status of the ten most frequently used American tests. In that review, she observed that very few researchers have addressed the issue of age norms. For example, it is only in recent years that norms for a wider range of age groups have been available for the widely-used

Wechsler Memory Scale. Lezak concluded that despite decades of awareness of the need for age norms for psychological (and neuropsychological) tests, few of the popularly used tests have adequate age norms. One notable exception in this regard has been a series of recent studies by Ivnik and colleagues (e.g., Ivnik, Malec, Tangalos, Peterson, Kokmen, & Kurland, 1990) who developed norms for elderly Americans for the Wechsler Adult Intelligence Scale - Revised (WAIS-R; Wechsler,1982), the Wechsler Memory Scale - Revised (WMS-R; Wechsler, 1987), and the Rey Auditory Verbal Learning Test (RAVLT; Rey,1964). A separate, but even more neglected issue, is that of the lack of local norms developed in New Zealand to which elderly New Zealanders might be compared. For these reasons the present study set out to develop some local norms for older New Zealanders on one widely used neuropsychological test, the Trail-Making Test.

The Trail-Making Test (TMT) is a popular test widely used by clinical psychologists, combining both brevity and ease of administration. It is economical (as it is in the public domain) and acceptable to most patients. According to Brown, Casey, Fisch and Neuringer (1958), the TMT originated from a prototype known as the Taylor Number Series. This test required the subject to draw lines connecting a series of numbers from 1-50 that were scattered randomly about a rectangular sheet of paper. Partington (1938, in Brown et al.) renamed and reconstructed it calling it the Test of Distributed Attention. It was then incorporated by Partington into the Army Individual Test of General Ability and renamed the Trail-Making Test. It subsequently became incorporated into the Halstead-Reitan battery (e.g., Reitan, 1955). The TMT was employed at that time chiefly as a short, simple screening device for the detection of brain damage. This reflects the prevailing belief of that era and the consequent concern with the detection of 'organicity". Subsequent theoretical and technical developments since then, especially modern imaging techniques, have meant that clinical neuropsychologists today are less concerned with diagnosing "organicity" per se, and more concerned with detailing the specific cognitive and behavioural deficits of patients and their correlates with identifiable lesions. Accordingly, the TMT is now regarded not merely as a screening measure for brain damage, but rather as a useful test of divided attention. In their authoritative work on the clinical neuropsychology of attention, van Zomeren and Brouwer (1995) describe Form B of the TMT as a measure of processing speed which demands mental flexibility and has elements of divided attention. They advocate its inclusion in any battery of tests concerned with attention and discuss it under the heading of divided attention (p.171).

Robertson, Ward, Ridgeway and Nimmo-Smith (1994) recently reported the results of a factor analysis which included numerous tests considered sensitive to attentional problems, including their own battery, and the TMT. They reported that the TMT (i.e., Trail B) loaded highly on a factor they called "visual selective attention/speed". They noted that the tests loading on this factor typically involved a considerable degree of selective attention and speed of processing. They described the common element of tests loading on that factor as "The ability to select target stimuli, while not selecting powerfully competing distractors" (p.10). Mirsky (1989) also reported a factor analytic study in which a battery of attentional tests were used. The

Table 1. Demographic Information on Participants by Age Group

	Age Range (years)				
	60–64	65–69	70–74	75–79	80+
Demographic Variable					
Female	14	15	13	15	17
Male	11	10	13	8	11
Mean Years of Education	11.6	10.9	11.0	10.7	10.5
Marital Status					
divorced	2	1	1	3	3
married	19	15	17	9	3
single	2	3	4	0	4
widowed	2	6	4	11	18
Residential Status					
independent	24	22	21	18	9
semi-independent	1	3	5	5	19

TMT (i.e., Trails B) loaded on a perceptual-motor speed factor which Mirsky viewed as measuring the *focusing* aspect of attention. Other tests loading highly on this factor included Talland Letter Cancellation, The Digit Symbol Substitution Test, and the Stroop Test.

In summary, the Trail-Making Test is a brief, easy to use, and economical test that is quite acceptable to most patients. Moreover, it provides important information on one specific dimension of cognitive functioning, namely attention (and more specifically visual selective attention). The aim of the present study was to compile a set of norms for the performance of elderly New Zealanders on the TMT. The data were collected by the second author (C.C.) who undertook this study as the research component of an M.A. (Applied) in Clinical and Community Psychology. Fuller details of the study and all raw data are contained in that thesis (Cavana, 1992).

Method

Participants

One hundred and twenty-seven Pakeha New Zealanders aged 60 or above participated in the study. Participants were grouped into five-year age brackets for normative purposes. Summary demographic data by age group are presented in Table 1.

Materials

In addition to Parts A and B of the TMT, all participants were administered the following measures:

Personal Information Questionnaire. This questionnaire was developed for the present study to provide background information on participants. It included questions on age, sex, current medication, general health, handedness, marital status, education, and residential status.

Geriatric Depression Scale. This is a 30-item questionnaire designed as a screening measure for depression among the elderly (Yesavage, Brink, Rose, & Adley, 1983). This measure was included because depression is considered to be relatively common among the elderly and also a possible (reversible) cause of cognitive impairment. Consequently, it was considered important to have some estimate of the frequency of depression among the group.

Vocabulary and Block Design subtests of the Wechsler Adult Intelligence Scale - Revised (Wechsler, 1981). These two scales provided an estimate of intellectual ability (Silverstein, 1982). This was considered important since several researchers have noted the influence of IQ upon TMT performance (e.g., Goul & Brown, 1970). A small problem arose here in that Silverstein (1982) only provides scaled scores for ages ranging up to 74 years. Hence it was decided to scale the raw scores for the two oldest age groups using Silverstein's norms for 70-74 year olds. This, at least, ensured a standardised and replicable procedure at the cost of underestimating slightly the correct scaled scores for the two older age groups.

Procedure

All participants were volunteers and were recruited from urban New Zealand areas (Auckland, Christchurch, Wellington and Dunedin). Participants were contacted through community groups, churches, social clubs, and through friends and acquaintances of the second author. All testing was conducted by the second author or by a colleague gathering data for a similar but separate study. Both testers were in the final year of a three-year training programme in clinical psychology.

The participants were clearly not a cross-section of the elderly population of New Zealand. This was beyond the scope of the present study. Rather, they were a large number of elderly people functioning independently or, at least, semi-independently in the community. All were fluent in English and reported no history of central nervous system disorders, major illnesses or uncorrected sensory deficits. Three participants from an initial sample of 130 were excluded on the basis of their extreme scores on some of the tests, leaving 127 participants.

Results

The mean score on the Geriatric Depression Scale was 5 with a standard deviation of 4. Scores ranged from 0 - 21 with a median of 4. Ten percent of participants had a score of 11 or greater indicating the presence of at least 'mild' depression. Estimated IQ's ranged from a minimum of 76 to a maximum of 144 with a mean of 108 (sd = 12.48).

The mean scores for each age group on both Trail A and Trail B are presented in Table 2. Inspection of Table 2 shows that scores for both A and B increase with increasing age with a rather steep increase apparent after age 79.

The correlations between the different measures administered are presented in Table 3. Inspection of Table 3 shows Trails A and B to be highly correlated ($r = .79$, $p<.0001$). Table 3 also shows moderate correlations between TMT B and both age ($r = .49$, $p<.0001$) and IQ ($r = -.54$, $p<.0001$). In view of these relationships it was decided to develop norms for the TMT that took into account both age and IQ using a regression model. The straightforward linear model showed signs of non-constant variance, especially for

Table 2 New Zealand Norms for the Trailmaking Test by Age Group

TRAIL A (in seconds)					
Age Group (years)	Sample Size	Mean	Standard Deviation	Minimum	Maximum
60–64	25	34.76	9.46	21	53
65–69	25	38.04	14.32	15	65
70–74	26	45.04	14.03	28	81
75–79	23	46.78	17.89	23	95
80+	28	60.21	24.38	28	110
TRAIL B (in seconds)					
Age Group (years)	Sample Size	Mean	Standard Deviation	Minimum	Maximum
60–64	25	74.96	23.67	43	131
65–69	25	94.56	50.90	29	290
70–74	26	123.69	57.72	55	246
75–79	23	127.13	86.02	50	480
80+	28	195.93	100.39	83	444

the higher age groups. We therefore sought transformations of the scores which would stabilise the variance. The transformation that best stabilised the variance was found to be a log transformation. The main regression equations are summarised in equations 1 and 2 below:

Equation 1: Log TRAIL A = 5.02 + 0.09 * AGE (group) - 0.014 * IQ

with the model $R^2(2, 127) = .381$, and both coefficients of the independent variables significant at the 1% level. No other variable (e.g., depression) added significantly to the explanatory power of the equation.

Equation 2: Log TRAIL B = 6.28 + 0.15 * AGE (group) - 0.02 IQ

with the model $R^2\ (2, 124) = .487$. No other variables were significant.

Table 3. Correlations Between Scores on TMT, IQ, GDS, Age and Education*

Pearson Correlation Coefficients

Measure	Trail A	Trail B	IQ	Age	GDS
Trail B	0.79				
IQ	-0.53	-0.54			
Age	0.46	0.50	-0.34		
GDS	0.20	0.16	-0.20	0.11	
Education (years)	-0.25	-0.27	0.40	-0.21	-0.14

*Note: n = 127
All correlations >0.16 significant at p<.05

The expected scores based upon these regression models, for Trail A by age group and IQ, are presented in Table 4. Similarly, in Table 5 expected scores by age and IQ for Trail B are shown. Inspection of both Tables 4 and 5 clearly demonstrates the effect of IQ upon TMT performance and highlights the need to take IQ, as well as age, into account when judging an individual's performance against their peers. Finally, norms are not presented separately for males and females as no significant differences (using t-tests) were observed between the sexes on either Trail A or B.

Discussion

The present study succeeded in providing age-stratified norms for New Zealanders aged above 60 years for the Trail-Making Test and highlights the importance of using age appropriate norms when making judgements about the performance of older people on neuropsychological tests. For example, in the present sample the mean time taken on Trail B by the 60 to 64-year-old group was approximately 75 seconds with a standard deviation of approximately 24. By contrast, the mean for 70 to 74-year-olds was almost two standard deviations higher at around 124 seconds. Hence, decisions based upon inappropriate age norms may yield quite faulty conclusions. One further aspect of the present data that was striking was the increase with age, not merely in absolute scores on the TMT, but also in the variability at each age band. Standard deviations increase almost fourfold across the five age groups. Thus, not only do the older participants perform more slowly on the TMT but with each increase in age band there is an increase in within-group variability. Again, this finding serves to confirm the importance of age specific norms.

The other major issue which emerged clearly in the present study was the importance of considering

Table 4. Expected Scores for Trail A by Age Group and IQ

(a) Expected Scores for Trail A (in seconds)

Age Group (years)	IQ Score 80	85	90	95	100	105	110	115	120	125	130
60–64	53	49	45	43	40	37	34	32	30	28	26
65–69	58	54	50	46	43	40	38	35	33	30	31
70–74	63	59	54	51	47	44	41	38	36	33	31
75–79	69	64	59	55	51	48	44	42	39	36	34
80+	75	70	65	61	56	52	49	45	42	39	37

(b) Upper 95% Confidence Limits for the Mean for Trail A (in seconds)

Age Group (years)	IQ Score 80	85	90	95	100	105	110	115	120	125	130
60–64	64	58	53	49	45	41	38	35	33	31	29
65–69	68	62	57	52	47	44	40	38	35	33	32
70–74	73	66	60	55	51	47	43	41	39	37	35
75–79	79	72	66	60	55	51	48	45	43	41	39
80+	86	79	73	67	62	58	54	51	48	46	43

an individual's intellectual ability when interpreting their score on the TMT (and presumably on many other cognitive tests as well). The data collected here demonstrated that even within a specific age range, people with quite different intellectual abilities will perform quite differently on the TMT. For example, the expected time for a 62-year-old with an estimated IQ of 80, to complete Trail B is 134 seconds (see Table 5). This can be compared with the expected time of 51 seconds for someone of the same age with an estimated IQ of 130. Such differences are great and failure to take both age and IQ into account when interpreting scores may yield grossly misleading conclusions. Although other writers have commented upon the necessity to take intellectual ability into account, we believe this is the first time that published

Table 5. Expected Scores for Trail B by Age Group and IQ

(a) Expected Scores for Trail B (in seconds)

Age Group (years)	IQ Score 80	85	90	95	100	105	110	115	120	125	130
60–64	134	122	111	101	91	83	75	69	62	57	51
65–69	156	142	129	117	107	97	88	80	73	66	60
70–74	182	165	150	137	124	113	103	93	85	77	70
75–79	212	193	175	159	145	131	119	109	99	90	81
80+	247	225	204	185	169	153	139	127	115	104	95

(b) Upper 95% Confidence Limits for the Mean for Trail B (in seconds)

Age Group (years)	IQ Score 80	85	90	95	100	105	110	115	120	125	130
60–64	169	150	133	118	106	95	85	77	71	65	60
65–69	191	169	150	133	119	106	96	87	80	74	69
70–74	218	193	171	151	135	121	110	101	93	87	81
75–79	251	223	198	176	158	143	130	120	111	103	96
80+	294	262	234	210	189	172	158	145	134	125	116

norms for elderly performance on the TMT have done so. At the same time it must be acknowledged that the scaled IQ scores for the two older age groups were estimated using 74-year-olds' norms. In this regard they almost certainly will represent underestimates of the "true" score. However, there seemed no better alternative given that appropriate age scaled scores were not available for the short form of the WAIS-R at the time these data were collected.

Readers intent on using the data reported here for patients over 74, but uncomfortable with this approach, have three options. Firstly, they may compare a specific patient with participants from Table 4 who have a slightly "higher" estimated IQ. Secondly, they can take a person's scores on the Vocabulary and Block Design subscales and estimate their IQ using the scaled scores given by Silverstein for 70 to 74-year-olds. As long as this score is calculated solely for comparison with the present normative group it should not violate any psychometric commandments. Thirdly, they can ignore intelligence completely and simply use the age related norms in Table 2.

There are a number of cautions, however, that should be clearly stated about the present norms. Firstly, the sample is not a representative cross-section of New Zealand elderly. All participants were volunteers, contacted through community groups and contacts of the senior author. In this regard they may reflect a chiefly middle-class sample. Certainly, the IQ data suggest they are slightly above average in intellectual ability. Moreover, the participants were entirely Pakeha New Zealanders. There were no Maori participants or representatives of other ethnic or racial groups. We would argue, however, that perhaps these groups merit their own norms rather than being represented by a small minority within the present sample. What we can say about the present sample is that it is a moderately large sample of older people who are living independently or semi-independently within the community.

Footnote

The authors wish to thank both Dr Bob Cavana, Senior Lecturer in Decision Sciences, and Dr Peter Smith, Institute of Statistics and Operations Research, at V.U.W., for help with the statistical analyses. We also thank Dr Gurusingham Sathiyandra of I.S.O.R. for advice on statistical programming. The comments of Michael Corballis and an anonymous reviewer also greatly improved our statistical analysis of the data.

References

Bak, J. S., & Greene, R. L. (1980). Changes in neuropsychological functioning in an aging population. *Journal of Consulting and Clinical Psychology, 4,* 395-399.

Bolla Wilson, K., & Bleecker, M. L. (1986). Influence of verbal intelligence, sex, age and education on the Rey Auditory Verbal Learning Test. *Developmental Neuropsychology, 2,* 203-211.

Bornstein, R. A. (1985). Normative data on selected neuropsychological measures from a non-clinical sample. *Journal of Clinical Psychology, 41* (5), 651-658.

Brown, E. C., Casey, A., Fisch, R. I., & Neuringer. (1958). Trailmaking Test as a screening device for the detection of brain damage. *Journal of Consulting Psychology, 22,* 469-474.

Cavana, C. M. (1992). *Norms for the Trailmaking Test in an older New Zealand population.* M.A.(Applied) in Clinical and Community Psychology thesis. Victoria University of Wellington.

Cipolotti, L., & Warrington, E. K. (1995). Neuropsychological assessment. *Journal of Neurology, Neurosurgery, and Psychiatry, 58,* 655-664.

Lock, S. (1978). Dementia - the quiet epidemic. Editorial. *British Medical Journal, 1,* 1-2.

Goul, W. R., & Brown, M. (1970). Effects of age and intelligence on Trailmaking Test performance and validity. *Perceptual and Motor Skills, 30,* 319-326.

Hart, S., & Semple, J. M. (1990). *Neuropsychology and the dementias.* Taylor & Francis Ltd., London.

Ivnik, R. J., Malec, J. F., Tangalos, E. G., Peterson, R. C., Kokmen, E., & Kurland, L. T. (1990). The Auditory Verbal Learning Test (AVLT) norms for 55 years and older. *Journal of Consulting and Clinical Psychology, 2,* 304-312.

Lezak, M. D. (1987). Norms for growing older. *Developmental Neuropsychology, 3* (1), 1-12.

Mirsky, A. F. (1989). The neuropsychology of attention: Elements of a complex behaviour. In E. Perecman (Ed.), *Interpreting theory and practice in clinical neuropsychology* (pp 75-91). Lawrence Erlbaum Associates, Hillsdale, N.J.

Prigatano, G. P. (1978). The Wechsler Memory Scale: A selective review of the literature. *Journal of Clinical Psychology, 34,* 816-832.

Reitan, R. M. (1955). The relation of the Trailmaking Test to organic brain damage. *Journal of Consulting Psychology, 19,* 393-394.

Rey, A. (1964). *L'examen clinique en psychologie.* Paris: Presses Universities de France.

Richmond, D., Baskett, J., Bonita, R., & Melding, P. (1995). *Care for older people in New Zealand: A report to the National Advisory Committee on Core Health and Disability Support Services.*

Robertson, I. M., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention Manual.* Thames Valley Test Company, Bury St Edmunds, England.

Silverstein, A. B. (1982). Two and four subtest short forms of the Wechsler Adult Intelligence Scale Revised. *Journal of Consulting and Clinical Psychology, 50* (3), 415-418.

van Zomeren, A. H., & Brouwer, W. H. (1994). *Clinical neuropsychology of attention.* Oxford University Press.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale - Revised.* New York: The Psychological Corporation.

Wechsler, D. (1987). *Manual for the Wechsler Memory Scale - Revised.* New York: Psychological Corporation.

Yesavage, J. A., Brink, T. L., Rose, T. L., & Adley, M. (1983). The geriatric depression rating scale: Comparison with other self report and psychiatric rating scales. In Crook, T., Ferris, S., & Bartus, R. (Eds.), *Assessment in geriatric psychopharmacology,* Connecticut: Mark Powley.

Address for correspondence:
Richard Siegert
School of Psychology
Victoria University of Wellington
PO Box 600
Wellington