Received: 2 October 2021 | Revised: 4 January 2022 | Accepted: 16 January 2022 | Published online: 28 February 2022

DOI: 10.1002/dad2.12292

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Digitally generated Trail Making Test data: Analysis using hidden Markov modeling

Mengtian Du[1,2] | Stacy L. Andersen[3] | Stephanie Cosentino[4,5] | Robert M. Boudreau[6] | Thomas T. Perls[3] | Paola Sebastiani[7]

[1] Department of Biostatistics, Boston University, Boston, Massachusetts, USA

[2] Analysis Group, 111 Huntington Ave. 14th floor, Boston, MA 02119, USA

[3] Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

[4] Department of Neurology and Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, New York, USA

[5] Gertrude H. Sergievsky Center, Columbia University, New York, New York, USA

[6] Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

[7] Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, Massachusetts, USA

Correspondence
Mengtian Du, Department of Biostatistics, Boston University, Boston, MA 02118, USA
E-mail: mandydu@bu.edu

Funding information
National Institute on Aging, Grant/Award Numbers: K01AG057798, 5U19AG063893, 5U01AG023749, 5U01AG023755, 5U01AG023712, 5U01AG023744, 5U01AG023746

Abstract

The Trail Making Test (TMT) is a neuropsychological test used to assess cognitive dysfunction. The TMT consists of two parts: TMT-A requires connecting numbers 1 to 25 sequentially; TMT-B requires connecting numbers 1 to 12 and letters A to L sequentially, alternating between numbers and letters. We propose using a digitally recorded version of TMT to capture cognitive or physical functions underlying test performance. We analyzed digital versions of TMT-A and -B to derive time metrics and used Bayesian hidden Markov models to extract additional metrics. We correlated these derived metrics with cognitive and physical function scores using regression. On both TMT-A and -B, digital metrics associated with graphomotor processing test scores and gait speed. Digital metrics on TMT-B were additionally associated with episodic memory test scores and grip strength. These metrics provide additional information of cognitive state and can differentiate cognitive and physical factors affecting test performance.

KEYWORDS

aging, Bayesian hidden Markov models, cognitive function, digital biomarker, Trail Making Test

1 | INTRODUCTION

The Trail Making Test (TMT) is one of the most commonly used and well-established neuropsychological tests for clinical evaluation of brain damage and diagnosis of age-related neurodegenerative diseases such as Alzheimer's disease. The original test was first introduced in the Army Individual Test Battery[1] as well as the Halstead-Reitan Neuropsychological Battery[2] in the 1940s. The version used in the current study[3] is a widely used paper-based version of the TMT, which consists of two parts. In TMT Part A (TMT-A), numbered circles are displayed on a piece of paper, and participants are instructed to use a pen to draw lines to connect the numbers in sequential order as quickly as they can. In TMT Part B (TMT-B), a series of numbers and letters are displayed and participants are instructed to connect the numbers and letters in alternate sequence (i.e., 1, A, 2, B, and so on). Traditionally, TMT scoring consists of total time to completion and the number of errors recorded by the examiner. The TMT was first used in hospitals as an indicator of brain damage.[4–6] Recent studies have shown that performance on the TMT-A is related to cognitive domains such as visual scanning, attention, and processing speed, while performance on TMT-B associates with more complex cognitive abilities including working memory, complex set maintenance, switching, and mental flexibility.[7–11]

Although the TMT has proven to be a highly sensitive test for detecting cognitive dysfunction, specific mechanisms underlying a well or poorly performed TMT are not differentiated in the overall time to completion. For example, a poorly scored TMT test might be a result of prolonged cognitive processing, motor difficulties associated with drawing, or both. The former implies impairment in cognitive abilities such as attention, working memory, cognitive flexibility, and visuoperception whereas the latter suggests dysfunction in physical abilities such as grip strength and dexterity. Studies have shown that collecting high-precision time-stamped data with the use of a digital pen during test completion and decomposing total time to completion into "thinking time" and "drawing time" provides added clinically useful information beyond standard scoring on the Clock Drawing Test[12–14] and the Digit Symbol Substitution Test.[15] These digital metrics have the potential to be informative about the relative contribution of cognitive and physical abilities to overall performance on the TMT as well.

In this study we used digital pen data from the TMT to decompose total time to completion and provide deeper insights into the cognitive or physical functions underlying overall performance on the TMT. We also propose a novel application of Bayesian hidden Markov models (HMMs) to perform automatic segmentations of the recorded drawings to create new digital metrics for the tests. Our hypothesis is that digitally recorded data streams provide additional information on cognitive state not captured by the overall time.

RESEARCH IN CONTEXT

1. **Systematic Review**: The Trail Making Test (TMT) is a well-established neuropsychological test. The authors used PubMed and meeting abstracts to accumulate knowledge related to the cognitive processes underlying the TMT and the processing of digital pen data from cognitive testing. These relevant citations are appropriately cited.
2. **Interpretation**: Our findings suggest that administration of the TMT with digital tools provides additional information about cognitive functions that is not captured by the overall time to completion.
3. **Future Directions**: Additional digital metrics could be extracted under the framework of Bayesian hidden Markov models. Examples include time required to switch between letter and number sequences and time spent while resting the pen in proximity of a target circle or in the process of drawing each connection. Future analysis could also investigate how these digital metrics associate with more comprehensive neuropsychological assessments of executive function as well as incident dementia.

2 | METHODS

2.1 | Study population and test measures

The Long Life Family Study (LLFS) is a multicenter longitudinal study of human longevity and healthy aging. The study has been described extensively in previous papers.[16,17] Briefly, families demonstrating clustering of longevity as measured by the relative survival probabilities of siblings in the proband generation were recruited to participate in the study.[17,18] All living siblings in the proband generation and their offspring and spouses were invited to participate in an in-home assessment of health and function and blood collection. To date, participants have completed up to two in-person assessments and are currently undergoing a third in-person assessment. Specific to the current analysis, which uses data from the second in-person assessment, participants completed a neuropsychological evaluation that included the Mini-Mental State Examination (MMSE), Hopkins Verbal Learning Test-Revised (HVLT-R), Logical Memory, verbal fluency (letter fluency for F, A, and S and category fluency for animals), Digit Symbol Substitution Test (DSST), Number Span Test, Clock Drawing Test, and TMT. Participants used a digital pen, described below, for all tests requiring written or drawn responses. Assessments of physical function included gait speed and grip strength to measure motor function and strength. Participants were followed annually (or every 3 years for those under 70 years of age) for updates in health and vital status and completed a modified version of the Telephone Interview for Cognitive Status (TICS-M) that sums the scores of questions Counting Backward, Word List Recall and Subtractions. TICS score was measured longitudinally with multiple measurements per participant, with mean follow-up time of 2.2 years; all other measures included in the analysis were cross-sectional from the second in-person visit.

An Anoto Live Ballpoint Pen (Model DP-201) was used to complete the TMT during the in-person visit. The digital pen looks like a wide-barrel ballpoint pen and writes in ink but also records and timestamps x and y coordinates across specially formatted paper 75 times per second, or approximately every 13 milliseconds. The recorded data stream is sectioned by each pen stroke, defined as a continuous drawing without lifting up the pen.

2.2 | Hidden Markov model and trail making tests

To extract information provided by the digital stream of coordinates, we implemented a novel way to perform automatic segmentation of the data using HMMs.[19] The intuition of the approach can be best described by considering the stream of coordinates in Figure 1A. Visually, one can identify the segments drawn to join the consecutive numbers 1 to 25. In machine learning, this segmentation can be

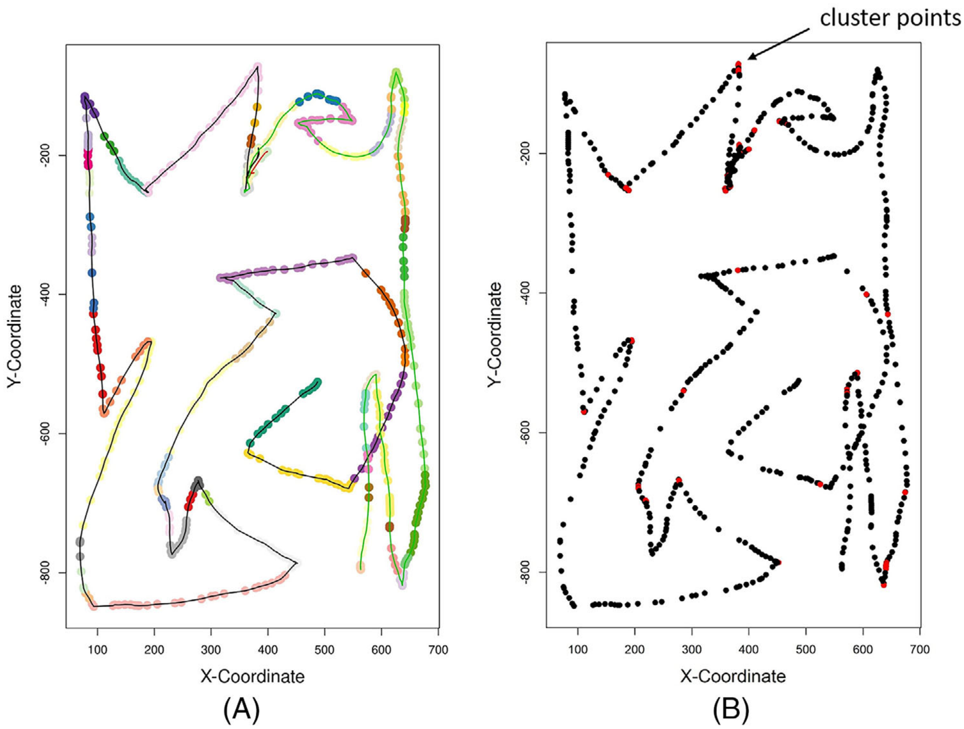

FIGURE 1 A, An example of a recreated drawing for the Trail Making Test Part A (TMT-A) from digital pen data after hidden Markov models segmentation; and (B) an example of drawing in TMT-A with cluster points in red

represented by an HMM, in which the stream of coordinates is modeled as a sequence of linear regressions and a series of hidden states that represent the various segments. A unique color corresponds to a HMM hidden state in Figure 1A. Details of the HMM method are described in the supporting information.

2.3 | Digital metrics

Using data from the digital pen, we extracted and derived several time metrics for TMT-A and TMT-B. We first defined a set of intuitive raw time variables, namely raw drawing time and raw non-drawing time. Raw drawing time was defined as the time spent while the digital pen was on the paper. Raw non-drawing time was defined as the time spent while the digital pen was lifted away from the paper, suggesting the participant was likely thinking or looking for the next number or letter in the sequence. A cluster of points was defined as a group of coordinates with pairwise distance $<\sqrt{2}$ coordinate units. Examples of clusters of points are illustrated in Figure 1B using the same TMT-A drawing as in Figure 1A, with clusters of points marked in red. A cluster of points indicates that the digital pen had moved less than one coordinate unit in both the vertical and horizontal directions, where one coordinate unit is equivalent to 0.3 millimeters.

Given that the time spent in these cluster coordinate pairs (i.e., while the pen is resting in place on the paper) as well as when the pen is lifted from the paper may reflect thinking or cognitive processing time rather than graphomotor speed, we then defined a set of derived time variables, namely derived drawing time and derived thinking time. Derived drawing time was calculated by subtracting the time spent in cluster points from the raw drawing time, and derived thinking time was calculated by adding time spent in cluster points to the raw non-drawing time. The derived time variables provide us a more accurate understanding of the decomposition of the total completion time in which derived drawing time captures the time the pen is moving across the paper and derived thinking time reflects the time the pen is not on the paper as well as the time the pen is resting in place on the paper. Using the results of the HMM segmentation, we extracted several metrics including number of segmentations and maximum length (in coordinate units) of the segments, for both TMT-A and TMT-B.

2.4 | Statistical analysis

We examined the association between these new derived metrics and raw scores of traditional metrics of cognitive function, which include TICS, DSST, category fluency for animals, Number Span Forward and Backward scores, and Logical Memory immediate and delayed recall, HVLT-R total recall, as well as metrics of physical function, including gait speed and grip strength using generalized estimating equations (GEE) with exchangeable correlation structure to account for family clustering. Additional traditional raw scores (e.g., HVLT delayed recall) were omitted to reduce multiple comparisons. The analyses were limited to the subset of participants who had successfully completed the tests (i.e., connected all dots in sequence within 5 minutes,

TABLE 1 Demographic characteristics and test scores of participants who completed the Trail Making Tests

	TMT-A	TMT-B	*P*-value (t-test)
N	2172	2014	
Age at visit 2 mean (SD), years	71.9 (11.1)	70.3 (9.7)	**<.001**
Sex, male (%)	978 (45%)	905 (44.9%)	.95
Education, college, and above (%)	1130 (52%)	1100 (54.6%)	.09
Test scores at Visit 2 (SD)			
TICS	15.7 (4.2)	16.1 (3.8)	**.01**
Animal Fluency	21.3 (6.5)	22 (6.1)	**<.001**
DSST	45.1 (13.8)	46.9 (12.4)	**<.001**
Number Span Forward	7.4 (2.3)	7.4 (2.3)	.46
Number Span Backward	6.3 (2)	6.4 (2)	**.049**
Logical Memory-Immediate	13.5 (4.3)	14 (4)	**<.001**
Logical Memory-Delayed	12 (4.8)	12.5 (4.5)	**<.001**
HVLT-R Total Recall	23.6 (6.2)	24.4 (5.6)	**<.001**
Gait Speed (m/s)	1.0 (0.3)	1.0 (0.2)	**<.001**
Grip Strength (kg)	27.9 (11)	28.9 (10.7)	**.003**

Abbreviations: DSST, Digit Symbol Substitution Test; HVLT-R, Hopkins Verbal Learning Test-Revised; SD, standard deviationTICS, Telephone Interview for Cognitive Status; TMT-A, Trail Making Test Part A; TMT-B, Trail Making Test Part B.
Note: Significant p-values (< 0.05) are displayed in bold.

including any errors that were corrected), because only these participants would get examiner-timed scores on their performances. The MMSE data were not included in this analysis because most of the study participants reached a perfect score. The GEE models adjusted for age at test, sex, education level, and familial longevity if significant, and included overall completion time and digital TMT metrics as additional predictors. An indicator variable "spouse" was created to take on value 1 if an individual was a spouse control, 0 if an individual was a member of a long-lived family. We first performed stepwise variable selection always keeping age, sex, and education level in the model. Using the selected variables, GEE models with exchangeable correlation structures were fitted accounting for within family correlations for all cross-sectional measures and accounting for within-subject and within-family correlations for the longitudinal TICS scores. We conducted these analyses separately in TMT-A and TMT-B. We retained variables with significance level <0.05, but will only discuss results with significance levels that pass the Bonferroni correction of multiple testing, which is 0.05/10 = 0.005.

The HMM analysis was conducted in R using the rjags package and the GEE analysis was implemented in SAS 9.4 using PROC HPGENSELECT for variable selection and PROC GENMOD for the final parameter estimates of GEE models.

3 | RESULTS

Out of 2778 LLFS participants who completed the second in-person visit with cognitive testing, 2364 participants attempted the TMT-A and 2172 successfully completed the test. Only 2330 participants attempted the TMT-B and 2014 successfully completed the test. Three hundred three tests were not included in the analysis due to loss of the digital pen files from improper storage by one of the field centers and did not count as participants who attempted the tests. Table 1 summarizes the demographic characteristics and test scores of the participants who were included in this analysis. Figure 2 and Figure 3 show the pairwise scatter plots of the digital metrics and completion time in TMT-A and TMT-B, respectively. Overall completion time and derived drawing time exhibit a strong correlation with $P = .85$ in TMT-A and $P = .73$ in TMT-B. A weak to moderate correlation is shown between overall completion time and number of HMM segments, with $P = .51$ in TMT-A and $P = .48$ in TMT-B.

Tables 2A and 2B and Tables S2a, S2b, S3a, and S3b in supporting information show the GEE parameter estimates of the analysis of the new metrics derived from the digital data from the TMT-A, while Tables 3A and 3B and Tables S4a, S4b, S5a, and S5b in supporting information show the GEE parameter estimates of the analysis of the new metrics derived from the digital data of the TMT-B. Overall, completion time was significantly associated with all cognitive and physical test measures when it was the only additional predictor in the GEE models. Derived drawing time was significantly associated with all test scores while the other digital metrics including derived thinking time, number of HMM segments, and maximum length of HMM segments were significantly associated with selected test scores. When adjusted for completion time, the digital metrics remained significant for selected test scores. Parameter estimates of the GEE models that included both completion time and digital metrics with significance levels that pass the Bonferroni correction for multiple testing are summarized below. Interpretation of results from GEE models that included only comple-

FIGURE 2 Pairwise scatter plot matrix for metrics in Trail Making Test Part A

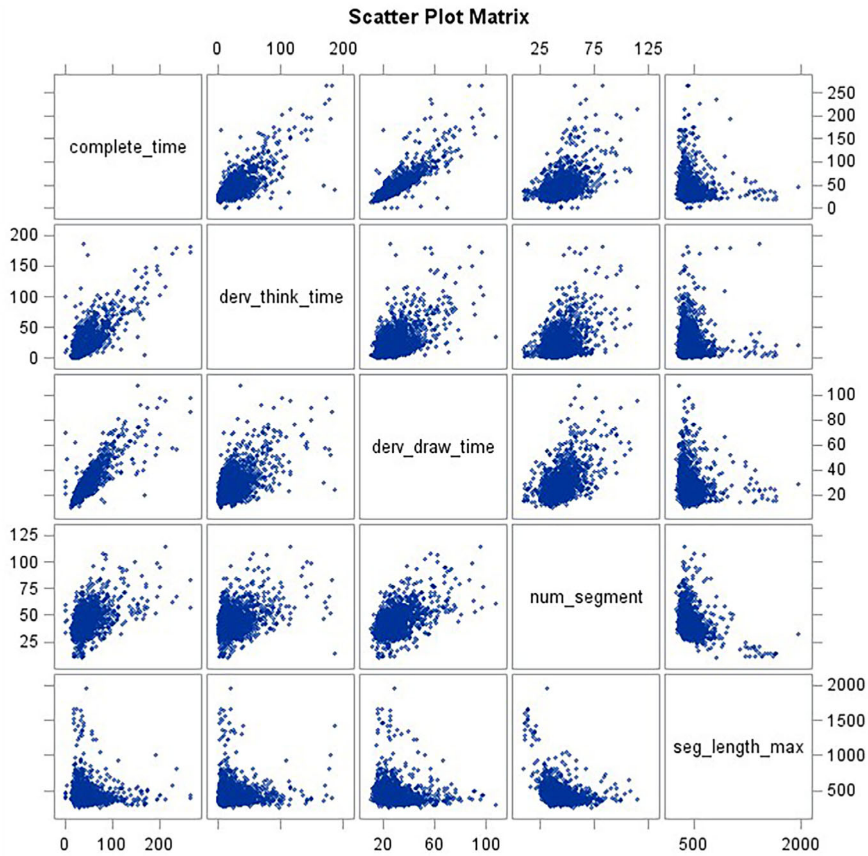

FIGURE 3 Pairwise scatter plot matrix for metrics in Trail Making Test Part B

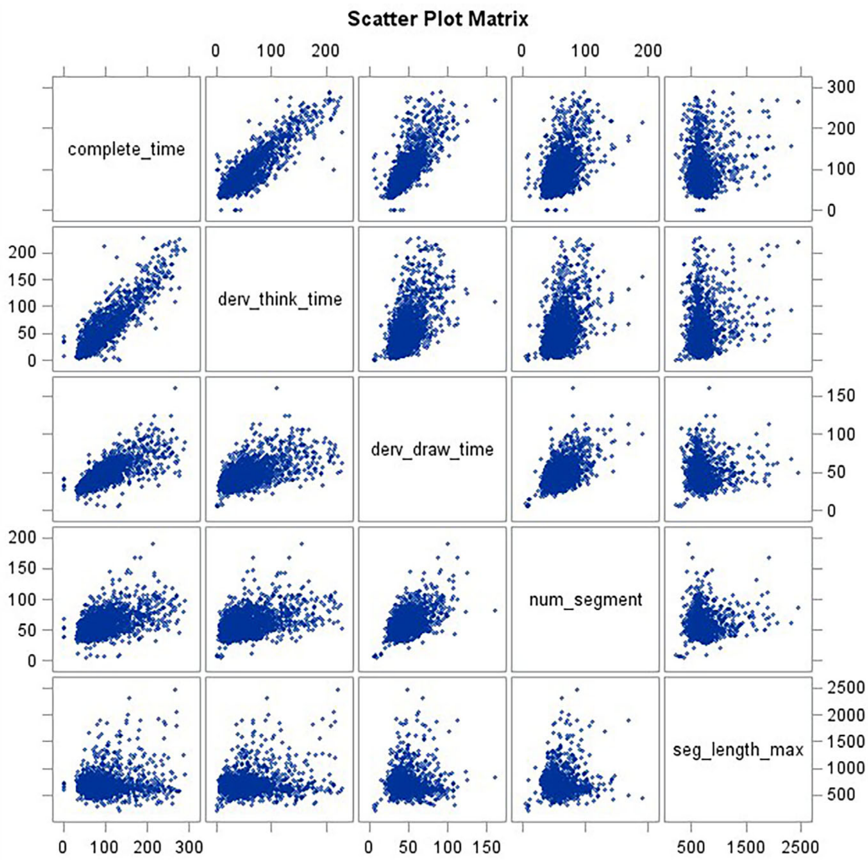

TABLE 2A Parameter estimates of GEE models using completion time and digital metrics as predictors, TMT-A

	TICS		DSST		Animal Fluency		Number Span–Forward		Number Span–Backward	
	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)
(Intercept)	23.20	<.0001	83.09	<.0001	34.60	<.0001	7.64	<.0001	6.40	<.0001
Age	−0.12	<.0001	−0.40	<.0001	−0.19	<.0001	−0.01	0.01	−0.02	0.0003
Sex (male)	−1.07	<.0001	−3.85	<.0001	−0.03	0.91	−0.01	0.95	−0.05	0.52
Education	0.28	<.0001	0.81	<.0001	0.26	<.0001	0.12	<.0001	0.15	<.0001
Spouse			−0.99	0.02						
Completion time	-0.05	<.0001	−0.08	0.0001	−0.05	<.0001	-0.01	<.0001	-0.01	<.0001
Derived drawing time			−0.33	<.0001						
Derived thinking time	0.02	0.001								
Number of segments			−0.10	<.0001	−0.03	0.03				
Maximum length of segments										

Abbreviations: DSST, Digit Symbol Substitution Test; GEE, generalized estimating equation; TICS, Telephone Interview for Cognitive Status; TMT-A, Trail Making Test Part A.

TABLE 2B Parameter estimates of GEE models using completion time and digital metrics as predictors, TMT-A

	Logical Memory–Immediate Recall		Logical Memory–Delayed Recall		HVLT-R Total Recall		Gait Speed		Grip Strength	
	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)
(Intercept)	18.16	<.0001	18.78	<.0001	35.25	<.0001	1.81	<.0001	52.93	<.0001
Age	−0.08	<.0001	−0.12	<.0001	−0.18	<.0001	−0.01	<.0001	−0.37	<.0001
Sex(male)	−0.86	<.0001	−1.12	<.0001	−2.68	<.0001	0.03	<.0001	14.45	<.0001
Education	0.24	<.0001	0.28	<.0001	0.38	<.0001	0.004	0.02	−0.17	0.001
Spouse										
Completion time	−0.02	0.004	−0.03	<.0001	−0.06	<.0001			−0.04	<.0001
Derived drawing time							−0.004	<.0001		
Derived thinking time	−0.02	0.01								
Number of segments							−0.002	0.001	−0.04	0.04
Maximum length of segments							−0.0001	0.02		

Abbreviations: GEE, generalized estimating equation; HVLT-R, Hopkins Verbal Learning Test-Revised; TMT-A, Trail Making Test Part A.

tion time or only digital metrics are available in the supporting information.

For TMT-A, in the model (Tables 2A and 2B) in which we included both completion time and digital metrics in the GEE models, completion time remained significantly associated with all test scores except gait speed. Derived drawing time was negatively associated with gait speed suggesting that the time spent in making connections between numbers had more significant associations with gait speed compared to the traditionally used completion time (parameter estimate = –0.004, standard deviation [SD]: 0.001, $P < .0001$). Derived drawing time was also significantly associated with the DSST score (parameter estimate = –0.33, SD: 0.04, $P < .0001$), in addition to completion time. This suggested the derived drawing time explained additional variance of the DSST score that completion time did not explain. Derived thinking time significantly predicted TICS score over follow-up but with a very small effect. Higher number of HMM segment was associated with lower DSST scores and slower gait speed. For one SD of additional segments (11.3), the DSST score was expected to decrease by 1.1 points (SD: 0.23, $P < .0001$).

For TMT-B, as shown in Tables 3A and 3B, in the model in which the GEE analyses included both completion time and digital metrics, completion time was negatively associated with all test scores except for grip strength, for which derived drawing time had a significant negative association in place of completion time.

TABLE 3A Parameter estimates of GEE models using completion time and digital metrics as predictors, TMT-B

	TICS		DSST		Animal Fluency		Number Span–Forward		Number Span–Backward	
	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)
(Intercept)	22.24	<.0001	74.19	<.0001	32.47	<.0001	7.51	<.0001	5.82	<.0001
Age	−0.08	<.0001	−0.30	<.0001	−0.14	<.0001	0.001	0.81	0.01	0.18
Sex(male)	−1.11	<.0001	−3.95	<.0001	−0.17	0.50	0.04	0.69	−0.04	0.64
Education	0.20	<.0001	0.66	<.0001	0.20	<.0001	0.09	<.0001	0.13	<.0001
Completion time	−0.02	<.0001	−0.08	<.0001	−0.03	<.0001	−0.01	<.0001	−0.02	<.0001
Derived drawing time	−0.02	0.03	−0.16	<.0001						
Derived thinking time										
Number of segments										
Maximum length of segments			0.004	<.0001						

Abbreviations: DSST, Digit Symbol Substitution Test; GEE, generalized estimating equation; TICS, Telephone Interview for Cognitive Status; TMT-A, Trail Making Test Part A.

TABLE 3B Parameter estimates of GEE models using completion time and digital metrics as predictors, TMT-B

	Logical Memory–Immediate Recall		Logical Memory Delayed Recall		HVLT-R Total Recall		Gait Speed		Grip Strength	
	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)	Estimate	Pr(> |Z|)
(Intercept)	17.33	<.0001	17.60	<.0001	32.21	<.0001	1.63	<.0001	52.60	<.0001
Age	−0.06	<.0001	−0.08	<.0001	−0.12	<.0001	−0.01	<.0001	−0.36	<.0001
Sex (male)	−0.82	<.0001	−1.08	<.0001	−2.77	<.0001	0.03	0.001	14.90	<.0001
Education	0.19	<.0001	0.23	<.0001	0.33	<.0001	0.005	0.01	−0.19	0.0002
Completion time	−0.02	<.0001	−0.02	<.0001	−0.02	<.0001	−0.001	<.0001		
Derived drawing time					−0.03	0.002			−0.06	<.0001
Derived thinking time										
Number of segments					0.01	0.04	−0.001	0.001		
Maximum length of segments					0.001	0.02				

Abbreviations: GEE, generalized estimating equation; HVLT-R, Hopkins Verbal Learning Test-Revised; TMT-A, Trail Making Test Part A.

Derived drawing time was also negatively associated with outcome measures of DSST (parameter estimate = –0.16, SD: 0.02, $P < .0001$) and HVLT-R Total Recall (parameter estimate = –0.03, SD: 0.01, $P = .002$). The number of HMM segments was associated with lower gait speed by 0.02 m/s (SD: 0.005, $P = .001$) for every 1 SD additional segments (17.9). Last, each coordinate unit SD increase (196.8) in maximum length of HMM segments was associated with higher score in DSST by 0.79 points (SD: 0.2, $P < .0001$).

4 | DISCUSSION

In this article we extracted metrics from digitally recorded TMTs by deriving time variables and using HMM to perform automatic segmentation of the recorded coordinates. We then analyzed the associations between these TMT metrics and other cognitive and physical function test scores. The overall results suggest that the digital metrics may provide additional information about underlying deficits in addition to the time used to complete the tests.

The analyses suggest that digital metrics of drawing time, thinking time, and number and length of HMM segments are associated with cognitive and physical functions. On TMT-A, drawing time and number of HMM segments were negatively associated with cognitive and physical outcome measures. Thinking time was not associated with any of the outcome measures when controlling for derived drawing time, which may point to the low cognitive processing demands of sequencing overlearned information (i.e., number sequences). Similarly, on TMT-B, drawing time and maximum length of HMM segments were associated with some of the cognitive and physical outcome measures but in contrast to TMT-A, thinking time was associated with cognitive function outcomes even beyond their associations with derived drawing time. The association of TMT-B thinking time with DSST, number span backward, and decline in TICS score over follow-up suggests that it is able to capture additional facets of psychomotor processing speed, auditory attention, and working memory that may be distinct from cognitive processing during the intervals between connecting the dots. In addition to differences in cognitive function, thinking time may also be capturing facets of personality and emotion that have been associated with executive function test performance, such as neuroticism[20] or timed-test–induced stress,[21] and should be investigated in future studies.

Based on these results, metrics from HMM segmentation provide additional information on cognitive and physical functions underlying performance on the TMT, which exceeds the overall completion time, the traditional metric of performance. On the TMT-A, the association of drawing time and number of segments with DSST and gait speed, even when controlling for overall completion time, points to the shared contribution of motor function across these tasks although shared effects of processing speed and attention may also be factors. Similar to TMT-A, the digital motor metrics (i.e., drawing time, maximum length of segments, and number of segments) on the TMT-B were associated with DSST and gait speed, above and beyond total time. However, TMT-B digital metrics were also associated with HVLT-R and grip strength. The association of digital metrics with HVLT-R on the TMT-B but not the TMT-A task condition points to underlying shared cognitive demands on learning and working memory that are specific to the number–letter sequencing task as has also been seen using traditional TMT scores.[22] Yet, the association of drawing time rather than thinking time with HVLT-R score may reflect remnants of thinking time remaining in the drawing time variable, or perhaps captures an aspect of learning that is exhibited during drawing. Longer maximum length of segments was associated with higher scores on the DSST, and thus may be an indicator of better processing speed and attention.

Several studies have used fully digital versions of the TMT, using iPads or Android-based applications on tablets or personal computers.[23–25] Dahmen et al.[26] and Fellows et al.[24] implemented a digital version of the TMT and extracted information such as pauses, pen lifts, time spent inside circles, and time between circles, with more sophisticated metrics extracted for TMT-B including average time before numbers or letters. The digital pen used in our study captures similar information such as pen location on the page and pressure. In our approach we chose to distinguish between writing time and drawing time as opposed to classifying the coordinate pairs as inside or outside of circles, or before or after letters, with the goal of isolating cognitive processing while not drawing. Additionally, the HMM segmentations provide information about detectable turn of direction in the drawing and allowed us to classify pauses (in the form of cluster coordinate pairs) as thinking time. Our innovative application of the HMM in this digital version of the TMT allows us to mathematically quantify and classify the recorded drawings in a holistic perspective.

Our findings of associations between digital metrics and neuropsychological test performance are also in line with those from versions of the TMT administered on a tablet. Fellows et al.[24] found that digital metrics from the TMT-A were associated with performance on the Symbol Digital Modalities Test, a processing speed test similar to the DSST which in our study was associated with several TMT metrics including drawing time and number/length of HMM segments on both conditions, as well as thinking time on TMT-B. Fellows et al.[24] also found that digital metrics on TMT-B were associated with tests of executive function including inhibitory control and visual working memory, a finding supported in our study wherein TMT-B thinking time was associated with number span backward. Despite broad similarities in results across studies, there are some differences in findings in our study (e.g., the association of processing speed with TMT-B metrics), likely reflecting differences in the algorithms used to create each digital metric and the cognitive processes that each metric captures (ie., pen lift duration vs. thinking time). Further research is needed to identify the underlying cognitive process(es) associated with each metric and to determine which algorithms are the most reliable. Additionally, there may be other digital metrics beyond those in the existing studies that relate to cognitive and personality outcome measures, such as thinking time after an error as an indication of ability to incorporate feedback and reestablish set, or changes in performance as the number of processed targets increases, which may relate to visual scanning abilities and susceptibility to interfering targets.

There may be clinical utility to the digital metrics derived in this study. The Boston Process Approach to neuropsychological assessment stresses the decomposition of cognitive test performance into its multifactorial components to better understand the underlying cognitive constructs and improve clinical decision making.[27] In an effort to isolate the executive component of performance on TMT-B from the visual scanning and graphomotor speed components, clinicians often subtract completion time on TMT-A from completion time on TMT-B. However, there are some biases in this method as the spatial array of dots on each test are not directly comparable; connections on TMT-B are longer and have more interfering stimuli than TMT-A.[28] In contrast, the digital metrics allow us to separate aspects of cognitive processing and motor function within each task, thereby reducing the effects of variations in the task stimuli and potential differences in testing conditions and test engagement between TMT-A and TMT-B, and offering the potential to eliminate administration of TMT-A.

One limitation of the current study is that drawing time may not be a pure metric of graphomotor function and may still contain facets of cognitive processing as suggested by the association of drawing time with other tests of cognitive function that do not require written responses (e.g., the HVLT-R). Participants may be thinking about their next move while making connections or may slow down but not completely stop drawing when trying to locate their intended target and thus these periods of cognitive processing would be included in the drawing time. The identification of these and other additional behavioral manifestations of cognitive processing would help to refine the drawing time and thinking time metrics. Second, the standardized administration of the TMT requires the examiner to stop the examinee whenever he or she commits an error so that it may be corrected. This artificially biases those with more errors to have longer thinking time; however, longer thinking time and a greater number of errors should both be associated with poorer cognitive function. Third, these analyses were performed only among individuals who were able to complete the test within the allotted 5 minutes thereby biasing the sample toward more cognitively healthy individuals and reducing the amount of variability in cognitive test scores. Finally, the only tests of executive function administered in this study were the DSST and number span backwards as well as verbal fluency, which in part draws on executive functions for optimal performance. More comprehensive assessments of executive function that include tests of set maintenance and switching may reveal additional associations with digital metrics, particularly for TMT-B.

5 | CONCLUSION

Digital technologies capture data on cognitive and physical components of test performance, in some cases even beyond what is captured by traditional test scores. Digital metrics derived from these data have the potential to provide added value to even brief cognitive assessments that may be used to better understand the relative contributions of specific cognitive and physical functions underlying test performance.

ACKNOWLEDGEMENTS

The authors were supported by the National Institute on Aging (K01AG057798 to S.L.A., 5U19AG063893 5U01AG023749 to S.C., 5U01AG023755 to T.T.P., 5U01AG023712, 5U01AG023744, 5U01AG023746); and the Marty and Paulette Samowitz Foundation to T.T.P. The funding organizations had no role in study design, data collection, analysis and interpretation of the data, the writing of the report, or in the decision to submit the article for publication.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTIONS

Mengtian Du and Stacy L Andersen: statistical analysis and drafted the manuscript; Paola Sebastiani: designed the study; Mengtian Du, Stephanie Cosentino, Robert M Boudreau, Thomas T. Perls, and Paola Sebastiani: interpretation of the results. All authors edited the manuscript.

REFERENCES

1. War Department, A.G.s.O., Washington, DC, Army individual test battery manual of directions and scoring. 1944.
2. Mazur-Mosiewicz A, Dean RS. Halstead-Reitan neuropsychological test battery. In: Goldstein S, Naglieri JA, eds. *Encyclopedia of Child Behavior and Development*. Springer US; 2011:727-731.
3. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills*. 1958; 8:271-276.
4. Reitan RM. The relation of the trail making test to organic brain damage. *J Consult Psychol*. 1955; 19(5):393-394.
5. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills*. 1958; 8(3):271-276.
6. Spreen O, Benton AL. Comparative studies of some psychological tests for cerebral damage. *J Nerv Ment Dis*. 1965:323-333.
7. Crowe SF. The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the trail making test. *J Clin Psychol*. 1998; 54(5): 585-591.
8. Sánchez-Cubillo I, Periáñez JA, Adrover-Roig D, et al. Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *J Int Neuropsychol Soc*. 2009; 15(3):438.
9. Arbuthnott K, Frank J. Trail making test, Part B as a measure of executive control: validation using a set-switching paradigm. *J Clin Exp Neuropsychol*. 2000; 22(4):518-528.
10. Oosterman JM, Vogels RLC, Van Harten B, et al. Assessing mental flexibility: neuroanatomical and neuropsychological correlates of the trail making test in elderly people. *Clin Neuropsychol*. 2010; 24(2):203-219.
11. Lezak MD. *Neuropsychological Assessment*. 5th ed.. Oxford University Press; 2012. xxv, 1161.
12. Davis R, Libon DJ, R Au, Pitman D, Penney DL. THink: inferring cognitive status from subtle behaviors. *Proc Conf AAAI Artif Intell*. 2014; 2014:2898-2905.
13. Libon DJ, Penney DL, Davis R, et al. Deficits in processing speed and decision making in relapsingremitting multiple sclerosis: the digit clock drawing test (dCDT). *J Multiple Scleros*. 2014; 1:113.
14. Davoudi A, Dion C, Amini S, et al. Classifying non-dementia and Alzheimer's disease/vascular dementia patients using kinematic, time-based, and visuospatial parameters: the digital clock drawing test. *J Alzheimers Dis*. 2021; 82:47-57.
15. Andersen SL, Sweigart B, Glynn NW, et al. Digital technology differentiates graphomotor and information processing speed patterns of behavior. *J Alzheimers Dis*. 2021; 82:17-32.
16. Newman AB, Glynn NW, Taylor CA, et al. Health and function of participants in the long life family study: a comparison with other cohorts. *Aging (Albany NY)*. 2011; 3(1):63-76.
17. Sebastiani P, Hadley EC, Province M, et al. A family longevity selection score: ranking sibships by their longevity, size, and availability for study. *Am J Epidemiol*. 2009; 170(12):1555-1562.
18. Wojczynski MK, Jiuan Lin S, Sebastiani P, et al. NIA long life family study: objectives, design, and heritability of cross-sectional and longitudinal phenotypes. *J Gerontol: Series A*. 2021.
19. Baum LE, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann Mathe Statis*. 1970; 41(1):164-171.
20. Murdock KW, Oddi KB, Bridgett DJ. Cognitive correlates of personality links between executive functioning and the big five personality traits. *J Individ Diff*. 2013; 34:97-104.
21. Starcke K, et al. Effects of acute laboratory stress on executive functions. *Front Psychol*. 2016; 7(461).
22. Vanderploeg RD, Schinka JA, Retzlaff P. Relationships between measures of auditory verbal learning and executive functioning. *J Clin Exp Neuropsychol*. 1994; 16(2):243-252.
23. Makizako H, Shimada H, Park H, et al. Evaluation of multidimensional neurocognitive function using a tablet personal computer: test-retest reliability and validity in community-dwelling older adults. *Geriatr Gerontol Int*. 2013; 13(4):860-866.
24. Fellows RP, Dahmen J, Cook D, Schmitter-Edgecombe M. Multicomponent analysis of a digital Trail Making Test. *Clin Neuropsychol*. 2017; 31(1):154-167.
25. Lunardini F, Luperto M, Daniele K, et al. Validity of digital Trail Making Test and Bells Test in elderlies. in 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2019.

26. Dahmen J, Cook D, Fellows R, Schmitter-Edgecombe M. An analysis of a digital variant of the trail making test using machine learning techniques. *Technol Health Care*. 2017; 25(2):251-264.
27. Ashendorf L, Swenson R, & Libon D. The Boston Process Approach to Neuropsychological Assessment: A Practitioner's Guide. The Boston Process Approach to Neuropsychological Assessment: A Practitioner's Guide. In: Ashendorf L, Swenson R, and Libon D, eds. *The Boston Process Approach to Neuropsychological Assessment: A Practitioner's Guide. The Boston Process Approach to Neuropsychological Assessment: A Practitioner's Guide*. Oxford University Press;2013:431-xxviii, 431.
28. Gaudino EA, Geisler MW, & Squires NK. Construct validity in the Trail Making Test: What makes Part B harder? *J Clin Exp Neuropsychol*. 1995;17(4):529–535. https://doi.org/10.1080/01688639508405143

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Du M, Andersen SL, Cosentino S, Boudreau RM, Perls TT, Sebastiani P. Digitally generated Trail Making Test data: Analysis using hidden Markov modeling. *Alzheimer's Dement*. 2022;14:e12292. https://doi.org/10.1002/dad2.12292