Digital Administration of the TMT: An Analysis of Intraindividual Variability in Computerized Testing

Dhanishka Pohuja

Master Thesis - Clinical Neuropsychology

S*3793117*
August 2022
Department of Psychology
University of Groningen
Examiner/Daily supervisor: Sarah Tol, MSc

Abstract

Neuropsychological testing is ever evolving into a new era of digitalization. This shift ensures newfound benefits for analyzing the progression of neurodegeneration since early prognosis. The purpose of the present study was to establish the interchangeability between a digital variant of the TMT and the traditional paper and pencil version. Moreso, whether the dTMT would be able to add to the traditional outcome measures, through the calculation of the MD, ISD, and rISD IIV measures, in order to be able to distinguish between those with or without an acquired brain injury, was of interest. Additionally, potential predictor variables which might have had an influence over the IIVs were assessed. Results showed that the dTMT was correlated with the pTMT and that the MD A, ISD A, and rISD A added value to the dTMT A and B, excluding the ratio score. Also, all part Bs of the IIV measures also added value to all variables of the dTMT and all three IIV measures were able to differentiate between ABI and HC participants. Age was a constant predictor for all of the MD and ISD IIVs in both the ABI and HC cohorts. Education level seemed to only have a predictor effect for the HC group and not for the ABI group for all but the rISD measures. CVA had a predictor effect for the dTMT B's MD, ISD, and rISD. This research was able to provide a stepping stone into how the dTMT has newfound potential into improving prognostic and diagnostic care through the application of IIV.

Keywords: digital TMT, intraindividual variability (IIV), trail making test, neuropsychological assessment, acquired brain injury

Digital Administration of the TMT: An Analysis of Intraindividual Variability in Computerized Testing

Neuropsychology includes the utilization of numerous methods and technologies that are applied to collect invaluable data for research, diagnosis, and treatment. Indubitably, the most used means are tests, questionnaires, and assessments (Muñiz & Bartram, 2007), which have found their origin from neuropsychological theories and procedures developed between 50 to 150 years ago (Bilder & Reise, 2018). Consequently, traditional (i.e., paper and pencil-based) testing does not aptly detect subtle cognitive changes, thus, computer-based testing has the potential to identify subtle impairments quicker than paper-based testing which in turn increases the chance of early impairment detection, diagnosis, and a more efficient prognosis (Bjorngrim et al., 2019). It has been noted that the differences between paper based and digital testing have become imperceptible as technology has become very advanced and common in daily use (Noyes & Garland, 2008; Bjorngrim et al., 2019). However, key differences between digital and paper and pencil-based testing include the advantage of the automatic generation of alternative forms for repetitive testing, uniformity in scoring, stimuli control, cost efficiency, feasibility of technological use, and the ability to live monitor and store the participant's movement and responses (Gualtieri & Johnson, 2006). This reduces the burden of standardized precision and accuracy from the clinician, and rather, more time to make behavioral observations which are important in diagnostics and rehabilitation. Thus, clinical psychological assessment based on paper and pencil assessments in the current age seems outdated, does not compare to current advances in cognitive research, which confines the capability for an accurate detection or diagnosis of mental illness or cognitive degeneration (Bilder & Reise, 2018).

Currently, many administered psychological tests and batteries are being performed and, scored, thus cognition is being assessed, through digitalization (Kane & Kay, 1992;

Parsey & Schmitter-Edgecombe, 2013) which consistently is showing a pattern of advantages over the paper and pencil assessments (Bauer et al., 2012; Cernich et al., 2007). This shift into digitalization ensures a uniformed approach to stimuli being presented, standardized scoring, data retrieval and storage, accurate response time stamps, and new beneficial possibilities which are unavailable through the traditional method of paper and pencil tests (Spreij et al., 2020). In addition to, research suggests that digital tests are more accessible, precise in assessment, and able to provide alternative forms (Gualtieri & Johnson, 2006). Therefore, the use of computerized assessments is currently being explored, as it is not only more cost-effective, rather it also reduces work-related burden of health care workers. Other advantages of using digitalized assessments include: (1) the capability to assess more individuals efficiently, (2) precise time stamp measurements on tasks, (3) ability to detect variability in performance (Dahmen et al., 2016), (4) standardization of tests, (5) automatic, instant scoring (Spreij et al., 2020), (6) performance and tests conveniently stored in online databases and patient's profiles, (7) automatic generation of cognitive reports, (8) administrating measures in various languages (Bauer et al., 2012), (9) time and cost effective compared to paper and pencil tests (de Vries et al., 2017) and lastly, (10) ethically and legally compliant measures (Cernich et al., 2007).

In light of the aforementioned evidence in favor of computerized assessments, de Vries et al., (2017) recognized the potential of using an electronic assessment battery in order to diagnose and test patients with acquired brain injury (ABI), especially those with perceptual disorders such as hemispatial neglect, spatial memory disorders, agnosias, and motor perceptual disorders, as they greatly influence life satisfaction. With the aim to improve the rehabilitation experience of ABI patients, the researchers conducted a Delphi study in order to produce an ideal standardized screening battery, known as the Diagnose Niet Aangeboren Hersenletsel (DiaNAH, translated Diagnosis of Acquired Brain Injury),

with screening for cognitive impairments especially with visual perception in mind. The research resulted in a panel of 17 international experts in the field of visual perception and ABIs creating a battery of eleven tests with a total administration time of 30 minutes. Of particular interest is one of the tests included in the DiaNAH Test Battery, the Trail Making Test (TMT). Historically, the trail making test was first introduced in the Army Individual Test Battery (1944) (Tombaugh 2004) by the U.S. Army in order to test general ability and intelligence, and then incorporated into the Halstead-Reitan Battery to measure brain and nervous system functioning (Tombaugh 2004; Reitan-Wolfson, 1986). Now, the TMT is frequently administered as a neuropsychological test that assess executive functioning (Park & Schott, 2021; Faria et al., 2015; Reitan, 1955), visual search, visual planning, visuomotor control, attention, and memory (Lin et al., 2021; Halstead, 1947; Reitan, 1971; Stuss et al., 2001; Strauss et al., 2006; Lezak et al., 2012). Therefore, the TMT is used to evaluate a plethora of neurological disorders and cerebral functioning. Moreover, the TMT comprises of two parts; TMT A and TMT B. The TMT A assesses cognitive abilities consisting of processing speed, visual search, and tracking (Fellows et al., 2016). On the other hand, the TMT B assesses mainly more demanding cognitive abilities such as divided attention (Kortte et al., 2002; Karimpoor et al., 2017), sequencing and working memory, which also correlates with the domains assessed in the WAIS-III Digit Span Backwards (Sánchez-Cubillo et al., 2009; Fellow et al., 2016). There is evidence of a correlation with the cognitive abilities tested in the Wisconsin Card Sorting Test (Chaytor et al., 2006; Kortte et al., 2002; Lamberty et al., 1994; Langenecker et al., 2007; O'Donnell et al., 1994; Ríos et al., 2004; Spikman, et al., 2001; Fellows et al., 2016), which corroborates the TMT B's association with set shifting (Fellows et al., 2016; Arbuthnott & Frank, 2000; Sánchez-Cubillo et al., 2009), executive control (Fellows et al., 2016), and flexibility (Strauss et al., 2006; Fellows et al., 2016), because of its task of alternating between letter and number stimuli.

Therefore, looking at the cognitive load associated with the TMT, especially part B, it is clear that the test has a sensitivity to neuropsychological decline, especially in those with mild cognitive impairment (Specka et al., 2021; Chapman et al., 2010; Ewers et al., 2012; Gomar et al., 2011), brain injury (Specka et al., 2021; Hanks et al., 2008), and stroke (Specka et al., 2021; Wiberg et al., 2012). Various cognitive domains are affected after a patient experiences a stroke, which also rely on stroke etiology, what parts of the brain were affected, time since onset, and age (Torrisi et al., 2019; Zucchella et al., 2014). In fact, Wiberg et al. (2010) observed that performance on the TMT B may be able to forecast future cerebrovascular accident (CVA) in older patients, aged 70+. One major consequence of stroke are visual field defects since it occurs in 8% to 31% of patients affected (Haan et al., 2016; Feigenson et al., 1977; Gilhotra et al., 2002). Generally, research shows that higher age or normal aging affected overall TMT performance; an increase in age decreased performance (Specka et al., 2021; Cavaco et al., 2013; Giovagnoli et al., 1996; Hashimoto et al., 2006; Perianez et al., 2007; Seo et al., 2006; Wecker et al., 2000). Research also indicates that other demographic data, such as education level, age, gender, and choice of handedness, might have an effect on TMT performance. Education level might influence performance on the TMT (Woods et al., 2015; Tombaugh, 2004; Fine et al., 2011) as education seems to have an impact on the TMT B more than the TMT A (Tombaugh, 2004), with those with a low educational level to be impacted the most (Giovagnoli et al., 1996; Amodio et al., 2002). Some studies report gender to have an effect on the TMT (Specka et al., 2021; Bezdicek et al., 2012), but it seems to have smaller effects than age or education level (Specka et al., 2021; Cangoz et al., 2009; Giovagnoli et al., 1996; Hester et al., 2005; Seo et al., 2006). Lastly, Bracken et al. (2019) reported an influence of handedness on both parts of the TMT as participants who were left-handed tended to take more time to complete the test.

Accordingly, the digitalized version of the TMT from the DiaNAH Test Battery, the dTMT, is meant to be a replica of Tests A and B of the TMT, thus, the underpinnings and cognitive demands and abilities used in the paper and pencil method, pTMT, has a possibility of being replicated within the dTMT as found in Bracken et al's. (2019) research. Thus, the aforementioned clinical and demographic variables that influence performance on the pTMT may subsequently have an effect on the dTMT. For example, research found that age (Park & Schott., 2021) and handedness (Bracken et al., 2019) to have an effect on the pTMT and their version of the dTMT, respectively. Furthermore, the present study's dTMT provides added valuable information in addition to the outcome measures from the pTMT. For instance, a plethora of data is collected from the current study's dTMT, such as time stamps between the subsequent stimuli, mistakes within the test, the instance when a participant lift their stylus or pauses during the test, real time recordings and images from the testing, and the ability to store a complete profile of the administration process of the test. This allows the clinician to have ample data to assess and relive the administration process, if necessary, in order to curate a patient report for diagnostic purposes, which cannot be achieved through the pTMT. These additional benefits were also corroborated in a study that also established that the TMT is a good indicator for many cognitive dysfunctions and the computerized version may provide better insight than the traditional method because of all of the novel data collected through the dTMT, which is not provided by the pTMT (Dahmen et al., 2016).

In addition to, there has been growing interest into further extrapolating the potential computerized testing has to offer within neuropsychological diagnostic and prognostic care, which has led to the capability to assess between tests or inter-stimuli data within tests to calculate the intraindividual variability (IIV) of a participant. IIV in neuropsychological testing is the examination of inconsistencies, fluctuations, and changes in a participant's performance within tests or between tests (Burton et al., 2002). Research specifies that

calculating IIV in the evaluation of cognitive functioning may be decisive, as it may be linked to neurological or behavioral degeneration (Hultsch et al., 2000). For example, a study determined that IIV could distinguish cerebral dysfunction caused by injury, disease, or aging, as it could envisage variability in behavior or cognition (Hultsch et al., 2000). In the domain of cognitive speed, IIV may be defined as the variation between trials of a reaction time assessment (Bielak & Anstey, 2019). In this case, those with neurological degenerative disorders such as dementia (Gorus et al., 2008; Bielak & Anstey, 2019), Parkinson's disease (de Frias et al., 2012; Bielak & Anstey, 2019), and traumatic brain injury (TBI) (Stuss et al., 2003; Bielak & Anstey, 2019), tend to score higher on IIVs than healthy controls (HC). Another research also claimed that IIV has promise to help neuropsychologists to understand and detect underlying cognitive issues and neurological development in individuals (Stawski et al., 2017). Remarkably, both the dTMT and intraindividual variability have the prognostic capability to indicate neurological degression and are affected by similar clinical and demographic variables, such as age and education level (Hultsch et al., 2000; Bielak & Anstey, 2019; Schretlen et al., 2003). Hence, calculating IIV values from the dTMT would be appealing in order to see whether removing the confounding effects of the aforementioned variables would refine the outcome measures obtained from the dTMT and to gauge their prognostic and diagnostic applicability within patients with ABI.

Moreover, multiple ways to calculate a measure for IIV have been proposed throughout literature, three of which will be studied in the current study; the Mean Difference (MD) (Schretlen et al., 2003), residualized Interstimulus Interval (rISD) (Hultsch et al., 2000; Bielak & Anstey, 2019), and the raw Interstimulus Interval (ISD) (Hultsch et al., 2000; Bielak & Anstey, 2019). In general, the time it took to connect each stimulus within the dTMT was crucial in the calculation of the IIVs, which is a novel way to calculate them. The MD has been previously calculated by Schretlen et al. (2003) in order to aid cerebral

dysfunction diagnostics through intraindividual variation. Research found that the MD scores increased with age. Interestingly, the researchers concluded that IIV is also common in 'normal adults,' and that variability alone should not be the sole reason for diagnostics. Comparatively, the current study will assess IIV as added value to traditional outcome measures gleaned from the pTMT, and as a supplementary addition to a patient's complete profile in order to ascertain prognosis and diagnosis. The rISD was calculated in a way that incorporated age, gender, and education level, in order to account for their confounding effects, and create a standardized IIV value. Meanwhile, the raw ISD has been discouraged from use for assessment of within and between person variability by previous research (Hultsch et al., 2008; Bielak & Anstey, 2019), however, due to slight variations in how the current study calculates the raw ISD for the dTMT, it was deemed appropriate to still explore the possibilities of the raw ISD and add to the literature of the raw ISD. By assessing IIVs through the dTMT, the research is trying to gauge whether performance on the test would also reflect in individual cognitive abilities related to the TMT A or B.

Therefore, the objective of this current research is to discover and examine the possibilities of the use and application of a computerized version of the TMT. Previous research on the dTMT have shown practicability on performance and administration of the test using large sample sizes (Spreij et al., 2020), but did not explore the diagnostic capabilities of the dTMT. Another study of the dTMT by Salthouse & Fristoe (1995) failed to reflect the potential of modernized testing as there was no automatic data collection, storage, or patient profiles, however, they were able to establish comparability of traditional outcome measures of the pTMT to the dTMT, providing a small stepping stone into the validation of the dTMT. Dahmen et al. (2017) conducted a study on a dTMT that aimed to establish comparability with the pTMT, and whether or not there were any added features of the dTMT that could distinguish between healthy and 'neurologic' participants. The study was able to

establish interchangeability between the pTMT and dTMT, as there were strong correlations found between the two, and their machine learning methods were able to distinguish between healthy and neurologic participants. However, Dahmen et al (2017) did not calculate newfound outcome measures like the IIV from their dTMT, and just established correlation between the pTMT and a version of the dTMT. Furthermore, other studies that included the dTMT were found to be establishing validity and reliability of their screening tools in very specific samples, for example, tailored from a Japanese population (Makizako et al., 2012; Onoda et al., 2013). As well experimented the dTMT is, the current study is manipulating the potential of the dTMT in ways that have been scarcely explored. This is anticipated to be achieved through the assessment, calculation, and examination of whether intraindividual variability, calculated potentially through participant MD, rISD, and/or the raw ISD, can indicate cognitive abnormalities in healthy control groups and participants with a history of ABI. Specifically, the current study aims to compare the digital and paper and pencil versions of the TMT in and its added potential through four research questions. The first research question analyzes whether there a distinction between the paper and pencil vs digital version of the TMT within the HC. The second research question asks if the IIV measures provide added value to the digital version of the TMT belonging to the HC cohort. Thirdly, whether participants with ABI show greater IIV than HC on the TMT IIV measures is explored. Lastly, the final research question asks what predictor variables influence IIV in ABI and HC cohorts. The HCs' predictor variables include choice of handedness, gender, age, and/or level of education of the participants and whether those variables have any effect on the administration on the digital version of the TMT and their respective IIV scores. On the other hand, in addition to the aforementioned predictors, the ABI participants' variables also included ABI etiology, binocular visual field abnormalities, which side of the brain was affected by the ABI, and the most recently occurred ABI. Through the examination of all

these research questions, the current research aims to provide insight into the digitalization of the TMT and the implications of doing so in neuropsychological use.

Methods

Participants

A total of 405 Dutch participants were recruited with their age ranging from 21 to 87 years old with a mean age of around 58 (M=58.19, SD=14.86). One hundred and ninety participants identified as male (46.91%) while the remaining 215 participants identified as female (53.09%). Furthermore, the participants were divided into two groups; the healthy control group (HC) and the acquired brain injury group (ABI). HC participants were recruited through advertisements in local newspapers, public establishments such as university buildings and libraries, social media, and by word of mouth. Inclusion criteria for the HC group were a minimum age of 18, absence of cognitive impairments, and a Mini Mental State Examination (MMSE) (Molloy & Standish, 1997) score of 24 or higher. HC participants also received a small monetary compensation for their contribution to the study. The collection of the HC data was approved by the Ethics Committee of the Psychology Department at RUG. Conversely, inclusion criteria for the ABI group included minimum age of 18 when the participant suffered from an ABI, an absence of comorbidities that could mimic the onset or symptoms of an ABI, and an intact vision with or without correction through glasses or contact lens (minimum of Snellen 0.5+). No compensation was provided to this group because they were clinical patients of Royal Dutch Visio where the DiaNAH Test Battery (de Vries et al., 2017) was implemented for a post-injury clinical assessment which included a neuropsychological evaluation as well. Patient data was collected under the ethical standards of the latest version of the Declaration of Helsinki and was subsequently approved by the Medical Ethical Review Board of the UMCG.

Moreover, 220 participants, 110 each from the ABI and HC groups, were matched one on one to control for the possible confounding effects of age, gender, and education level. This 'Match' variable was used in the third research question whether the ABI group was scoring higher IIV scores than the HC group. Age was categorized into a range consisting of ten years (e.g., 40-49 years) except for a younger cohort of 18-29 years of age and those participants older than 80+. Educational level was established with the guidelines of the Dutch education rating scale system consisting of seven levels. 'Low education level' consisted of levels 1 through 4 (Verhage 1-4), 'Middle education level' was level 5 (Verhage 5), and a 'High educational level' was levels 6 through 7 (Verhage 6-7). Table 1 summarizes the demographic and clinical information for all three groups.

Table 1

Descriptive Statistics & Clinical Characteristics for Each Participant Group

	HC	ABI	Match
1. Gender (*N*)	230	175	220
Male (*N*) (%)	90 (39.13)	100 (57.14)	104 (47.27)
Female (*N*) (%)	140 (60.87)	75 (42.86)	116 (52.73)
2. Age (*M*, *SD*)	55.36, 16.07	61.91, 12.17	60.65, 13.16
3. Education Level (*N*)	229	175	220
Low (*N*)	15	49	20
Middle (*N*)	64	66	92
High (*N*)	151	60	108
4. Handedness (*N*)	230	173	220
Right (*N*) (%)	196 (85.22)	160 (92.48)	193 (87.73)
Left (*N*) (%)	27 (11.74)	11 (6.36)	21 (9.55)
Ambidextrous (*N*) (%)	7 (3.04)	2 (1.16)	6 (2.73)
5. Binocular VA (*N*)	0	175	110
0.5 – 1.00 (*N*) (%)		97 (55.43)	60 (54.55)
1.05 – 1.60 (*N*) (%)		78 (44.57)	50 (45.45)
6. Binocular VF (*N*)	0	145	91
Normal (*N*) (%)		14 (9.66)	11 (12.09)
Doubtful (*N*) (%)		15 (10.34)	8 (8.79)
Impaired (*N*) (%)		114 (78.62)	71 (78.02)
Unreliable (*N*) (%)		2 (1.38)	1 (1.10)
7. Time Since Last ABI (Range (*M*, *SD*))		0 – 280 (10.21, 25.30)	0 – 74 (8.05, 9.89)

	HC	ABI	Match
8. ABI Etiology (*N*)	0	175	110
Stroke NOS (*N*) (%)		19 (10.86)	8 (7.27)
iCVA (*N*) (%)		136 (77.72)	87 (79.09)
hCVA (*N*) (%)		17 (9.71)	13 (11.82)
i-hCVA (*N*) (%)		3 (1.71)	2 (1.82)
9. Side of Stroke (*N*)	0	144	89
Left (*N*) (%)		55 (38.19)	34 (38.20)
Right (*N*) (%)		71 (49.31)	41 (46.07)
Both (*N*) (%)		18 (12.5)	14 (15.73)

Note. VA = visual acuity; VF = visual field; NOS = Not Otherwise Specified; iCVA = ischemic stroke; hCVA = hemorrhagic stroke; i-hCVA = both ischemic and hemorrhagic strokes. Time since the last recorded ABI was measured in months. In variables where the *N* or % is less than the group's total *N* or %, the data was unknown or not collected for that cohort.

Materials

The current research used data collected by neuropsychologists of a visual rehabilitation center, and research assistants of the University of Groningen implementing the DiaNAH Test Battery. For the purposes of the current research, only the data collected for the Trail Making Test portion of the battery was utilized and analyzed.

DiaNAH Test Battery

Every participant from the HC and ABI groups was administered the DiaNAH Test Battery (de Vries et al., 2017) which consisted of 11 appropriate tests that can measure impairments in visual perception (de Vries et al., 2017), especially in those with acquired brain injury. The eleven tests in the battery included those that were frequently used in neuropsychological testing in traditional paper and pencil format, such as the Trail Making Test, as well as new computerized tests, such as the Birthday Party Test (de Vries et al., 2020). The paper and pencil formatted tests were all converted into a digital version with Metrisquare B.V. hardware and DiagnoseIS software. This was administered on a Wacom 24-inch screen tablet, provided with an electronic stylus pen in order to mimic a pencil and/or pen. At the end of assessment, there was a possibility to generate a digital report, either

clinical or scientific, where the raw scores for each individual test, copies of the completed tests, demographic information, etc. were automatically compiled into a digital file. In the clinical report, all the raw test scores were provided, and if applicable, normed scores were also calculated. The scientific report could have been converted into an Excel or SPSS dataset for probable data manipulation. Currently, there is no normative data for the complete DiaNAH Test Battery (de Vries et al., 2017), however, the University of Groningen have been implementing the battery in a large group of participants without ABI in order to facilitate normative data collection and accumulate information on the validity of the DiaNAH. The HC dataset used in this TMT study was taken from the aforementioned data compilation.

Trail Making Test (TMT)

The Trail Making Test (TMT) was chosen to assess visuomotor speed in the DiaNAH Test Battery (de Vries et al., 2017). The ABI group contributed to dTMT data whereas the HC group took both the paper and pencil TMT (pTMT) and the dTMT versions. Both the pTMT and the dTMT consisted of Part A and Part B like the traditional administration of the neuropsychological test. Similarly, both versions of the test consisted of the same instructions, with the neuropsychologist or research assistant following the administrative guidelines of the traditional TMT. In Part A, participants were asked to link randomly dispersed stimuli containing numbers, from 1 through 25, in chronological order as quickly and efficiently as they could have. Likewise, in Part B, participants were asked to connect the randomly dispersed stimuli containing numbers and letters chronologically, however, alternating between number and letter, starting with number one first (e.g., 1 to A, A to 2, 2 to B, B to 3, etc.). In both parts, they had to do this without breaking the connecting lines and without lifting their stylus, pen, or pencil after initiating the task. In the event of an error, such as breaking the chronological order or missing a link, and if the participant did not

notice the error and correct it themselves, the administrator of the test should correct the participant immediately. For example, if the participant missed a link or connected the wrong pair of stimuli, then the administrator should point out the mistake and tell them which pair of stimuli should be linked instead, cross out the error line, and briefly remind them of the instructions. In the dTMT, when an error is made by the participant, a sound is emitted from the tablet, in which then the experimenter should quickly point out the error and remind the participant of the instructions. Also, in the dTMT, various time stamps were recorded automatically; the time between the start and finish of the task and the time it took between connecting one stimulus to the other. In the pTMT, only start and end time of each part was manually recorded with a stopwatch. Time was not stopped when an error was made or when the participant was being corrected in both versions. Scores were calculated with the same process in both versions; the outcome measures were: total time, derived ratio scores calculated as Part B/Part A, and inter-stimulus intervals (ISIs) calculated through the amount of time, in seconds, it took to connect a line from one circle to the next. Normative information is automatically generated from the raw scores from the dTMT, however, is not available for use as of yet as data is still being collected. Additional derived scores were calculated using the various time stamps collected by the dTMT, the intraindividual variability (IIV) scores.

Intraindividual Variability

The intraindividual variability was solely calculated from the dTMT as the calculation needed inter-stimulus intervals (ISIs) which were estimated from the time, in seconds, it took a participant to draw a line from one circle to the next. The ISIs were further refined to correct for the distance by measuring the space between each pair of stimuli in centimeters and then dividing each ISI with the distance in centimeters. By doing so, it was possible to create three different derived scores reflecting IIV, mainly, the Mean Difference (MD)

(Schretlen et al., 2003), residualized Interstimulus Interval (rISD) (Hultsch et al., 2000; Bielak & Anstey, 2019), and the raw Interstimulus Interval (ISD) (Hultsch et al., 2000; Bielak & Anstey, 2019). The MD score was calculated by subtracting the lowest ISI from the highest ISI. This was completed for both parts of the dTMT, for every HC and ABI participant, resulting in two MD scores per participant in the dataset, MD A and MD B. In addition to, the rISD was calculated through a regression analysis with variables age, gender, and educational level, which was dummy coded into 'middle' and 'high' as variables with 'low' as the reference variable, for every ISI. Therefore, there were 48 residuals for every participant for both parts of the dTMT; 24 residuals for Part A and 24 residuals for Part B. The 24 residuals for each part were then computed into a standard deviation (SD) for each participant, resulting in a SD for Part A and one for Part B. The raw ISD was retrieved by calculating the standard deviation across all of the ISIs.

Procedure

Each participant of the HC and ABI groups experienced the same initial intake process before the beginning of the experimental conditions. The participant entered a private room where only the participant and experimenter were present for the assessments. Then, the HC participants were debriefed about what the purpose of the study was and their privacy and anonymity were ensured. Afterward, the HC participants were asked to fill out a demographic questionnaire, and the Mini Mental State Examination (Molloy & Standish, 1997), in order to ensure the participant was eligible to continue as a HC. Since ABI participants were already diagnosed with a brain injury prior to being referred to the Dutch Royal Visio, the debrief and consent process was different from the HC, as the process was already incorporated into their clinical care, and were asked for permission to use their patient profile for research purposes. Both group participants were also reminded to wear their glasses or contact lens for corrected vision, and the distance of the tablet was adjusted to

conform to test requirements and/or their comfort. The participant was then asked to free draw on the tablet in order to adjust to the digital stylus as well as to check whether the stylus needed to be recalibrated. Participants of both experimenter groups completed the DiaNAH Test Battery (de Vries et al., 2017) in its fixed order.

Consecutively, participants completed the DiaNAH Test Battery (de Vries et al., 2017) The instructions for each of the 11 tests in the battery was shown on the experimenter's screen and it also reflected what the participant's movements were for each prompt, as the experimenter's screen showed the exact movements the participant made on their tablet. For example, when the participant drew a line from circle 1 to circle 2 on the dTMT A, the experimenter was able to see this in real time from their own tablet. Further, the total time and timeframe between each pair of stimuli are automatically recorded with time being recorded as soon as the participant touched the tablet with the stylus. Errors were automatically recorded; however, it was indicated by a double time log, thus, the experimenter also needed to pay attention to the errors manually.

Additionally, as part of a validation study, the HC group took all the tests or parallel versions of tests involved in the DiaNAH Test Battery (de Vries et al., 2017) in the form of the traditional paper and pencil method as well. Participants in this group always started with the digital administration before the paper and pencil version. Thus, the HC group partook in the tests twice, completing both forms subsequently one after the other with an optional 10-minute break in between. The total time of assessment for the HC group was approximately 2.5 hours. During the pTMT, the experimenter had to follow the traditional testing instructions and guidelines, manually take note of the start and end times for both parts, and physically point out errors and cross out error lines, if necessary. Scoring and psychometric information, such as norms, reliability, and validity of the traditional TMT was readily available as well.

Statistical Analyses

The dataset was analyzed using the 27th version of the software IBM SPSS Statistics. The data was examined for outliers, normality, linearity, and other assumptions as required by the specific analysis applied in each research question. In this case, assumptions for Paired Sample T-Tests, Spearman Correlations, Independent Sample T-Tests, and Multiple Regression were tested (as outlined below). Bonferroni post hoc corrections were applied to each of the research questions except for the Spearman Correlation analysis in the second research question.

RQ 1: Paired Sample T-Tests to Assess Differences Between the dTMT and pTMT. Paired Sample T-Tests were performed for the first research question in order to assess if there was a distinguishable difference between the pTMT and the dTMT within the HC population. The first pair was between the total time of the dTMT A and the total time of the pTMT, the second pair analyzed the total time of the dTMT B and the total time of the pTMT B, and the last pair differentiated between the ratio scores of the dTMT B/A and the pTMT B/A. Cohen's *d* was calculated for each Paired Samples T-Test in order to measure the effect size between the means. The interpretations of the effect sizes of Cohen (1988) were the benchmark used for analysis. The assumptions tested were: use of continuous (i.e., interval or ratio) dependent variables, the use of the same sample in both pairs, random sampling of the data from the population, the difference of the pairs having a normal distribution, and the absence of outliers in the difference between the pairs.

RQ 2: Correlation Between IIV Measures and dTMT Variables. Next, a descriptive and correlational analysis was conducted in order to gauge whether or not IIV measures, namely the MD, ISD, and rISD, had a relationship strong enough to be able to provide any added value to the outcome measures commonly calculated from the TMT. Thus, the IIV measures were compared with the dTMT A, dTMT B, and dTMT B/A ratio scores.

The HC cohort were the only participants used for the correlational analysis because the comparability and added value of the IIV measures were being assessed, and it had to be conducted without the influence of confounding ABI variability. as First, the assumptions for a Pearson's correlation analysis were tested. Thus, variables were computed at a continuous level, linear relationships between the variable were checked, significant outliers were assessed, and the variables involved were checked to see if they were normally distributed. There were violations found, thus, a Spearman's rho analysis replaced the Pearson's correlation (see Results section).

RQ 3: Comparison of IIV Measures Between the HC and ABI Groups. Thirdly, an Independent Sample T-test was computed in order to distinguish the difference in IIV measures between the HC and ABI groups and to assess whether ABI participants score more in the IIV measures than the HC participants. This was evaluated through two separate tests; one with only the 'Match' sample and only the non-standardized IIV measures (MD A, MD B, ISD A, and ISD B) and another test with all participants and only the standardized IIV measure, rISD A and rISD B. Just as the first research question, Cohen's *d* was used to interpret effect size of the t-tests. There were five assumptions tested for this analysis. The first checked whether the dependent variables were measured on a continuous scale and then the dependent variables were analyzed for normal distribution. Then, the independent variables were made sure to include two independent, categorical groups for analysis. After, the absence of outliers, independence of observations, and the test of homogeneity of variances were assessed.

RQ 4: Multiple Regression Analyses to Analyze Which Predictor Variables Influence IIV in HC and ABI Participants. The last research question consisted of separate multiple regression analyses, one with only the HC participants and another with the ABI participants. The HC group tested all the IIV measures with predictor variables handedness,

gender, age, and educational level, in order to see whether there is an influence from these variables on the IIV measures. Subsequently, all the ABI participants were tested with all the IIV measures with handedness, age, education level, gender, the etiology of the ABI, on what side the ABI was found, time since the last ABI, and the affected binocular visual field as predictor variables in order to gauge whether any of those variables had an influence on any of the IIV measures. Etiology of the ABIs included: Stroke NOS (Not Otherwise Specified, hCVA (hemorrhagic cerebrovascular accident), iCVA (ischemic cerebrovascular accident), and i-hCVA (both ischemic and hemorrhagic cerebrovascular accident). Localization of the ABI were either found in both, left, or right sides of the brain. Variables of the binocular visual field predictor included normal, doubtful, and unreliable vision. For these analyses, there were six assumptions analyzed for the multiple regressions. The first assumption checked the linearity between the independent and dependent variables. Then, the absence of multicollinearity was gauged through the analysis of collinearity, tolerance, and VIF scores. Residuals were computed and checked for independence through the Durbin Watson value, the variance of the residuals was made sure to be constant, and the normality of the residuals were also assessed. Finally, Cook's Distance was computed in order to make sure there were no influential data affecting our multiple regression models.

Results

General Comments on the Assumptions

Whenever the normality assumption was tested and failed to meet the requirement within the various statistical analyses used, the present study was able to overlook the violation of normality because of the large sample size collected in this current research, as it falls within the basis of the Central Limit Theorem (CLT) (Kwak & Kim, 2017). In theory, the CLT states that if the sample size of a study is large enough, then the means of the samples obtained through random sampling should be normally distributed with the mean and

variance, with no influence from the population distribution (Kwak & Kim, 2017). The CLT advises a sample size of 30 or greater as appropriate for the theorem to be upheld which means that the current research was able to utilize the CLT in its analyses as the sample size consisted of 405 participants. Thus, the current study was more than qualified to apply the CLT on the normality violations within each analysis and continue to make use of parametric testing as much as possible in order to upkeep statistical power and deduce more accurate conclusions from the data. Furthermore, the present study also did not remove any of the outliers from the data after careful consideration, as none of the outliers seemed non-representative of the overall population, which do tend to include extreme scores, or indicated a source of error. It is acknowledged that the influence of outliers is substantial on the outcomes of some statistical testing, thus, in those instances, the use of nonparametric tests was favored, such as the use of a Spearman Correlation in the second research question. Nonetheless, outliers were preserved in order to avoid statistical bias, type I errors, and to represent overall generalizability of a regular population (Gress et al., 2018). Subsequently, all tables and figures for the assumption testing for each research question could be found in Appendix A.

RQ 1: Is there a distinction between the paper and pencil vs digital version of the TMT within the HC?

Assumption Testing

Examination of the differences between the three Paired-Sample T-Tests for the first hypothesis demonstrated that there were violations in the normality assumption (Table A1). All the other assumptions of the Paired Sample-T-Test were satisfied appropriately; this included independent sampling of observations, dependent variables being measured on an incremental value, and the dependent variables were paired (Appendix A1-A6).

Statistical Analysis Results

Firstly, a Bonferroni correction was calculated because multiple Paired Sample T-Tests were conducted, adjusting the p-value to approximately 0.0167, in order to avoid the risk of type I errors (Kwak & Kim, 2017). There was a significant difference between the dTMT A (M=30.96, SD=10.58) and pTMT A (M=28.91, SD=11.17) with $t(227) = 3.82$, $p < 0.001$, d=0.25. This output suggested that there is variation in completion time between the dTMT A and the pTMT A and that participants took significantly more time, in seconds, to complete the dTMT A compared to the pTMT A. For the second Paired Sample T-Test, the results showed that there was also a significant difference between the dTMT B (M=67.47, SD=31.13) and pTMT B (M=58.73, SD=25.87); $t(226) = 6.20$ $p < 0.001$, d=0.41. The second test also suggested a contrast in completion time between the dTMT B and pTMT B with the HC group taking longer on the dTMT. The last Paired Sample T-Test's output exhibited a slightly non-significant difference between the ratio B/A scores of the dTMT (M=2.22, SD=0.67) and pTMT (M=2.09, SD=0.64) with $t(226) = 2.38$, p=0.018, d=0.16. This outcome may suggest that there were little dissimilarities between the derived scores between the dTMT and pTMT, but completion time remained longer in the dTMT (M=2.23, SD=0.67) compared to the pTMT (M=2.09, SD=0.64). Overall, the results suggested that participants of the HC group needed more time to complete the digital test than the paper test. Table 2 depicts the descriptive information for all the dTMT and pTMT variables used.

Table 2

Descriptive Statistics of the Digital and Paper TMT Variables Within the HC Cohort

	N	*M*	*SD*	Minimum	Maximum
dTMT A	230	31.01	10.58	13.0	76.7
dTMT B	230	68.41	32.32	24.50	234.60
dTMT B/A	230	2.23	0.67	0.90	4.66
pTMT A	228	28.91	11.17	11.00	88.00
pTMT B	227	58.73	25.87	0.00	156.00
pTMT B/A	227	2.09	0.64	0.00	5.30

Note. The variables of the dTMT and pTMT are total times in seconds. The 'B/A' variables for each of the TMT versions are a ratio score calculated by the division of part B by part A total times.

RQ 2: Do the IIV Measures Provide Added Value to the digital version of the TMT?

Assumption Testing

Examination of the normality outputs for the second hypothesis using Pearson's Correlation also demonstrated violations of the normality assumption for all variables tested (Table A2). Nevertheless, the assumptions of linearity, homoscedascity, and the absence of outliers were also violated (Appendix A7 to A60), thus, the use of a nonparametric test, specifically the Spearman Correlation, was appropriate in this case.

Statistical Analysis Results

Spearman's rho correlation coefficient was used to assess the relationship among both the parts of the dTMT and the IIV measures. Table 3 demonstrates the Spearman's intercorrelation amongst the variables. The interpretations for the analysis of a correlation coefficient defined by Schober et al. (2018) were used in this paper. The output showed that the dTMT A had a strong, significant, and positive correlation with the dTMT B scores and the dTMT A's MD, ISD, and rISD IIV measures. The dTMT A had a moderate, significant, and positive relationship with the dTMT B's MD, ISD, and rISD IIV measures. The dTMT A had a negative, non-significant, and negligible correlation with the derived score of the dTMT. Next, the dTMT B seemed to have a moderate, positive, and significant correlation with the derived score of the dTMT, and dTMT A's MD and rISD, while a strong, positive, and significant relationship with the dTMT B's MD, ISD, and rISD scores. The dTMT B also had a strong, positive, and significant correlation with the ISD score of part A. Lastly, for the derived score of the dTMT, there was a negative, weak, and nonsignificant correlation with the part A of the MD score. There was a significant, moderate, and positive correlation between the TMT B/A and the dTMT B's MD, ISD, and rISD IIV measures. The output showed a nonsignificant, positive, but negligible and weak relationships between the TMT B/A score and the dTMT A's ISD and rISD scores, respectively. Overall, there seems to be a

trend where the part A of all three IIV measures add value to the dTMT's A and B, but not the TMT B/A ratio score, while the part B IIV measures seem to add value to all three dTMT variables.

Table 3

Correlations Amongst the dTMT's Parts A & B Total Times and IIV Measures from the HC

	TMT A	TMT B	B/A	MD A	MD B	ISD A	ISD B	rISD A	rISD B
TMT A	--								
TMT B	0.709**	--							
B/A	-0.192**	0.491**	--						
MD A	0.658**	0.432**	-0.160*	--					
MD B	0.414**	0.757**	0.531**	0.314**	--				
ISD A	0.741**	0.496**	-0.169*	0.819**	0.335**	--			
ISD B	0.421**	0.803**	0.597**	0.318**	0.860**	0.378**	--		
rISD A	0.669**	0.473**	-0.126	0.891**	0.342**	0.874**	0.357**	--	
rISD B	0.420**	0.784**	0.565**	0.313**	0.935**	0.351**	0.923**	0.357**	--

Note. The above table depicts the Spearman rho between the dTMTs' A and B's completion time in seconds while the B/A variable is the ratio of the score of the TMT B divided by the TMT A score. Total *N* was 230 for all variables except for the rISD A and B, which was *N* = 229. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

RQ 3: Do the ABI group show greater IIV than the HC group on the TMT IIV measures?

Assumption Testing

For the third research question, two Independent Sample T-Tests were conducted; one for all of the non-standardized IIV measures, which include the MD and the ISD for the Match group, and another for the standardized IIV measure, rISD, for the whole sample. The assumption of independence was met; however, the assumption of normality of the dependent variables for each ABI and HC groups was violated (Table A3). Levene's Test for Equality of Variances demonstrated that the MD, ISD, and rISD were violating the homogeneity of variances assumption (Table A4). Thus, in order to compare the IIV measures between the HC and ABI groups, the 'Equal variances not assumed' portion of the Independent Sample T-Test was used for both occasions.

Statistical Analysis Results

Independent Samples T-Tests were utilized in order to deduce whether the ABI group score higher on the IIV measures than the HC group. The first test only including participants matched one on one with those with similar age, gender, and education level characteristics within their group (Match variable) and the all the MD and ISD variables. There was a significant difference in the MD A scores between the ABI participants (M=1.24, SD=1.28) and HC participants (M=0.77, SD=0.66); t(162.99) = 3.42, p = <0.001, d=0.46. Likewise, there was a significant difference in the MD B scores; ABI (M=3.37, SD=4.96) and HC participants (M=1.52, SD=1.32); t(124.30) = 3.79, p = <0.001, d=0.51. These results suggest that the ABI group did score higher than the HC group for the MD IIV measure. Next, there was a noticeable significant difference in the ISD A scores for the ABI group (M=8.92, SD=7.20) and the HC group (M=4.42, SD=2.95) with t(144.51) = 6.07, p = <0.001, d=0.82. For the ISD B scores, there was a significant difference in scores as well; ABI participants (M=8.84, SD=9.15) while the HC participants (M=3.91, SD=3.05) with t(132.94) = 5.36, p = <0.001, d=0.72. Again, the brain injured participants had higher scores than the healthy control group for the ISD IIV measures. These results suggest that, indeed, there is a difference in IIV scores between both groups as the participants with ABI tend to score higher than the HC group, even without the effects of possible confounding predictor variables that were controlled for in the present study.

Thereafter, the second test used the whole sample population with only the rISD IIV measures. There was a significant difference in the rISD A scores between the ABI (M=0.31, SD=0.27) and the HC (M=0.13, SD=0.10); with t(210.38) = 8.16, p = <0.001, d=0.91. Similarly, there was a significance in the difference between the rISD B scores between the ABI (M=0.78, SD=0.87) and the HC (M=0.27, SD=0.23); with t(190.71) = 7.51, p = <0.001, d=0.85. The output from the second Independent Samples T-Test suggests that there is a

differentiation between the scores of the two groups as the ABI participants scored higher than the healthy participants. The results from both the tests insinuate that all the IIV measures, MD, ISD, and rISD, are able to distinguish between those brain injured and healthy controls.

RQ4: What predictor variables influence IIV in the ABI and HC groups?

Assumption Testing

The fourth research question contained many predictors which were analyzed using the Multiple Regression method, one for each HC and ABI group, in order to understand which predictor variables, if any, had influence on the IIV measures within the HC or the ABI group. All of the IIV measures were log transformed in order to correct for the violation of each model's assumption of linearity, independence of residuals, homoscedascity of residuals, and the normality of residuals. The assumption of linearity between the independent and dependent variables was not met for all variables (see Appendix A). The second assumption of no multicollinearity in the data was met for all the variables. The third assumption of the residuals being independent was also met for both analyses. The fourth assumption, the homoscedascity of residuals, was also successfully established for all of the variables. The fifth assumption assumes that the residuals are normally distributed, which was non-violated for all the variables. The last assumption, lack of influential outliers, was met for all the variables.

Statistical Analysis Results

Multiple regression models were conducted in order to assess whether any predictor variables were able to influence the IIV scores for any of the measures. The first set of multiple regression analyses utilized only the healthy controls and analyzed variables age, gender, level of education, and handedness as predictors of all the IIV measures. A Bonferroni correction was applied, by dividing the alpha value by the number of predictors,

readjusting the p-value to 0.0083 for the MD and ISD measures, and p-value of 0.025 for the rISD measure, respectively. As seen in Table 4, in reference to the MD IIV measure, MD A and MD B had statistically significant regression equations. 'Age' was the only significant predictor for MD A, while both 'Age' and 'Education Level' were significant predictors for MD B. The table also shows that both ISD A and ISD B models were found to be significant. The significant predictor for ISD A was 'Age' while both 'Age' and 'Education Level' were significant predictors for ISD B. Lastly, Age, Gender, and Education Level were removed as predictors for the rISD IIV measure because they were already included in the calculation of the rISD values. Both the rISD A and rISD B multiple regression outputs were non-significant since neither of the choice of handedness variables were a significant predictor in either model.

Table 4

Results of the Multiple Regression Analyses for All the IIV Measures Within the HC Group

	t	*p*	*β*	*F (df)*	*p*	*Adj.* R^2
MD A						
Model				6.77 (6, 223)	<0.001	0.13
Age	6.07	<0.001	0.01			
Gender	-1.19	0.237	-0.05			
Edu Level Low	-0.12	0.903	-0.01			
Edu Level High	-0.24	0.814	-0.01			
Left-Handed	0.99	0.325	0.07			
Ambidextrous	0.12	0.970	0.02			
MD B						
Model				13.75 (6, 223)	<0.001	0.25
Age	6.49	<0.001	0.01			
Gender	-1.64	0.103	-0.06			
Edu Level Low	2.33	0.021	0.20			
Edu Level High	-2.82	0.005	-0.12			
Left-Handed	1.80	0.073	0.11			
Ambidextrous	-0.75	0.455	-0.08			
ISD A						
Model				8.26 (6, 223)	<0.001	0.16
Age	5.97	<0.001	0.01			
Gender	-1.78	0.076	-0.05			

	t	*p*	*β*	*F (df)*	*p*	*Adj.* R^2
Edu Level Low	0.71	0.482	0.05			
Edu Level High	-1.85	0.066	-0.06			
Left-Handed	0.23	0.819	0.01			
Ambidextrous	-0.63	0.531	-0.05			
ISD B						
Model				13.86 (6, 223)	<0.001	0.27
Age	6.50	<0.001	0.01			
Gender	-1.47	0.142	-0.05			
Edu Level Low	2.35	0.020	0.16			
Edu Level High	-3.21	0.002	-0.11			
Left-Handed	0.83	0.409	0.04			
Ambidextrous	-0.94	0.350	-0.08			
rISD A						
Model				0.96 (2, 226)	0.386	0.00
Left-Handed	0.15	0.881	0.01			
Ambidextrous	1.38	0.169	0.14			
rISD B						
Model				1.30 (2, 226)	0.274	0.00
Left-Handed	1.57	0.12	0.09			
Ambidextrous	0.50	0.62	0.05			

Note. Bolded variables indicate significant predictors of the associated model. Education was shortened to 'Edu.' Variables 'Left-Handed' and 'Ambidextrous' refer to the choice of handedness for each participant.

The second set of multiple regression analyses utilized only the ABI group and analyzed variables age, level of education, gender, handedness, the etiology of the ABI, the side the lesion was found, the binocular visual field affected, and time since the last ABI occurred as predictors of all the IIV measures. A Bonferroni correction was calculated by dividing the alpha value from the number of predictors, readjusting the p-value to 0.0033 for the MD and ISD measures and a p-value of 0.0045 for the rISD measures, respectively. Table 5 depicts the MD A and MD B regression models. Both models were statistically significant; 'Age' was the only significant predictor for the MD A measure while 'Age', 'Gender', and 'Stroke Etiology' were the significant predictors for the MD B measure, making the remaining predictors nonsignificant.

Table 5

Results of the Multiple Regression Analyses for the MD IIV Measure Within the ABI Group

	t	*p*	*β*	*F (df)*	*p*	*Adj.* R^2
MD A						
Model				2.43 (15, 151)	0.003	0.11
Age	4.57	<0.001	0.10			
Gender	-1.86	0.066	-0.11			
Edu Level Low	0.75	0.454	0.05			
Edu Level High	0.40	0.694	0.03			
Left-Handed	0.01	0.991	0.00			
Ambidextrous	-0.88	0.383	-0.23			
Recent ABI	1.17	0.244	0.00			
Stroke NOS	-0.33	0.739	-0.03			
hCVA	1.42	0.158	0.14			
i-hCVA	-0.11	0.912	-0.02			
Both Stroke	0.47	0.640	0.04			
Left Stroke	-0.35	0.724	-0.02			
Normal	-1.53	0.128	-0.16			
Doubtful	-0.32	0.750	-0.03			
Unreliable	0.04	0.967	0.01			
MD B						
Model				3.97 (15, 151)	<0.001	0.21
Age	4.15	<0.001	0.01			
Gender	-3.21	0.002	-0.17			
Edu Level Low	0.68	0.498	0.05			
Edu Level High	-1.13	0.262	-0.07			
Left-Handed	-0.84	0.402	-0.10			
Ambidextrous	-0.88	0.380	-0.22			
Recent ABI	1.78	0.078	0.00			
Stroke NOS	-0.69	0.489	-0.06			
hCVA	3.81	<0.001	0.34			
i-hCVA	1.37	0.172	0.27			
Both Stroke	1.35	0.179	0.12			
Left Stroke	2.02	0.045	0.12			
Normal	-1.54	0.127	-0.15			
Doubtful	-0.26	0.796	-0.03			
Unreliable	0.59	0.554	0.15			

Note. Bolded variables indicate significant predictors of the associated model. Education was shortened to 'Edu.' Variables 'Left-Handed' and 'Ambidextrous' refer to the choice of handedness. 'Recent ABI' is the time since last ABI. 'Stroke NOS', 'hCVA', and 'i-hCVA' are etiologies of stroke. 'Both Stroke' and 'Left Stroke' refer to stroke location. 'Normal', 'Doubtful', and 'Unreliable' refer to the effect on binocular visual fields.

Subsequently, the multiple regression models for the ISD A and ISD B measures showed statistical significance. The ISD A model depicted in Table 6 shows that 'Age' and binocular visual field were the significant predictors while the other predictors in the model were nonsignificant. In the ISD B model, 'Age', 'Gender', and 'Stroke Etiology' were the significant predictors while the other predictors remained nonsignificant.

Table 6

Results of the Multiple Regression Analyses for the ISD IIV Measure Within the ABI Group

	t	*p*	*β*	*F (df)*	*p*	*Adj.* R^2
ISD A						
Model				2.72 (15, 151)	<0.001	0.11
Age	4.39	<0.001	0.01			
Gender	-0.35	0.724	-0.12			
Edu Level Low	1.09	0.276	0.06			
Edu Level High	0.70	0.483	0.04			
Left-Handed	0.41	0.686	0.04			
Ambidextrous	-0.54	0.591	-0.11			
Recent ABI	-0.29	0.771	0.00			
Stroke NOS	-1.25	0.215	-0.10			
hCVA	1.72	0.088	0.13			
i-hCVA	0.75	0.456	0.13			
Both Stroke	-0.26	0.795	-0.02			
Left Stroke	-0.80	0.427	-0.04			
Normal	-2.89	0.004	-0.24			
Doubtful	-0.70	0.486	-0.06			
Unreliable	-0.60	0.553	-0.13			
ISD B						
Model				3.42 (15, 151)	<0.001	0.18
Age	4.06	<0.001	0.01			
Gender	-2.96	0.004	-0.14			
Edu Level Low	1.62	0.107	0.10			
Edu Level High	-0.40	0.689	-0.02			
Left-Handed	-0.75	0.456	-0.07			
Ambidextrous	-0.94	0.350	-0.21			
Recent ABI	0.65	0.520	0.00			
Stroke NOS	-1.03	0.303	-0.09			
hCVA	3.06	0.003	0.25			
i-hCVA	1.31	0.191	0.24			
Both Stroke	0.70	0.487	0.06			

	t	*p*	*β*	*F (df)*	*p*	*Adj.* R^2
Left Stroke	0.88	0.381	0.05			
Normal	-1.98	0.050	-0.17			
Doubtful	-0.66	0.513	-0.06			
Unreliable	0.94	0.351	0.21			

Note. Bolded variables indicate significant predictors of the associated model. Education was shortened to 'Edu.' Variables 'Left-Handed' and 'Ambidextrous' refer to the choice of handedness. 'Recent ABI' is the time since last ABI. 'Stroke NOS', 'hCVA', and 'i-hCVA' are etiologies of stroke. 'Both Stroke' and 'Left Stroke' refer to stroke location. 'Normal', 'Doubtful', and 'Unreliable' refer to the effect on binocular visual fields.

Finally, both the rISD A and rISD B models were statistically nonsignificant. Table 7 shows that the rISD A measure had no significant predictors, while the rISD B measure had 'Stroke Etiology' as a sole significant predictor.

Table 7

Results of the Multiple Regression Analyses for the ISD IIV Measure Within the ABI Group

	t	*p*	*β*	*F (df)*	*p*	*Adj.* R^2
rISD A						
Model				0.76 (11, 154)	0.677	-0.02
Left-Handed	0.11	0.910	0.01			
Ambidextrous	-0.20	0.840	-0.04			
Recent ABI	0.70	0.483	0.01			
Stroke NOS	-0.52	0.603	-0.04			
hCVA	1.05	0.295	0.08			
i_hCVA	0.03	0.980	0.00			
Both Stroke	0.13	0.897	0.01			
Left Stroke	-0.66	0.512	-0.04			
Normal	-2.30	0.023	-0.19			
Doubtful	-0.89	0.374	-0.08			
Unreliable	-0.53	0.598	-0.11			
rISD B						
Model				1.89 (11, 154)	0.045	0.12
Left-Handed	-0.95	0.344	-0.09			
Ambidextrous	-0.59	0.559	-0.13			
Recent ABI	1.38	0.168	0.00			
Stroke NOS	-0.56	0.577	-0.05			
hCVA	2.98	0.003	0.24			
i_hCVA	1.11	0.270	0.20			
Both Stroke	0.70	0.486	0.06			

	t	*p*	*β*	*F (df)*	*p*	*Adj.* R^2
Left Stroke	0.94	0.347	0.05			
Normal	-2.28	0.024	-0.20			
Doubtful	-0.71	0.480	-0.06			
Unreliable	0.60	0.551	0.13			

Note. Bolded variables indicate significant predictors of the associated model. Education was shortened to 'Edu.' Variables 'Left-Handed' and 'Ambidextrous' refer to the choice of handedness. 'Recent ABI' is the time since last ABI. 'Stroke NOS', 'hCVA', and 'i-hCVA' are etiologies of stroke. 'Both Stroke' and 'Left Stroke' refer to stroke location. 'Normal', 'Doubtful', and 'Unreliable' refer to the effect on binocular visual fields.

Discussion

The field of neuropsychology has been evolving within the last few decades with the adoption of technology within their assessment testing compared to the traditional hands-on approach. There has been sparked interest in the potential of digitalized testing and its prospective in revolutionizing the way diagnostics, treatment, and rehabilitation can be improved through added benefits available due to the nature of computerization. As such, this current study tried to extract the inventive potential of digitalizing the TMT. This was done by assessing the comparability of the pTMT and dTMT, and whether the dTMT is able to provide additional information onto traditional outcome measures of the pTMT that could help improve the future of diagnostic care in neuropsychology. This was conducted through the investigation of four research questions; the first question tried to establish comparability and differences between the pTMT and dTMT, in order to prove the core features found in the pTMT to be preserved and interchangeable in the dTMT. After establishing equivalency of the dTMT in comparison to the pTMT, the second research question examined whether the IIV measures calculated from the dTMT provided any additional information of value, in adjunct to the traditional outcome measures conceivable from the pTMT. The third question assessed whether the new outcome measures calculated from the dTMT, the IIV scores, are able to distinguish the participants with ABI from the healthy participants, and whether those injured scored higher than the HCs. Lastly, after confirming that ABI participants have higher

IIV scores compared to the HCs, the final research question delved a bit deeper into the IIV measures and tried to identify potential predictor variables which had any discernable influence over the IIV scores.

RQ 1: Is there a distinction between the paper and pencil vs digital version of the TMT within the HC?

Primarily, the initial question explored in the present study was whether the digital version of the TMT was comparable to the commonly used paper and pencil version of the TMT. According to the results obtained, there is a strong correlation between two pairs; the dTMT A and pTMT A and the dTMT B and pTMT B. This would indicate that the digital and paper versions of the TMT are indeed comparable, and that both parts of the dTMT assess neuropsychological functioning similarly to both parts of the pTMT. Similar studies reported high correlations with their paper and digital versions of the TMT as well, further supporting the current study's analysis (Bracken et al., 2018; Dahmen et al., 2017; Lunardini et al., 2019; Park & Schott, 2021). Subsequently, the derived ratio scores of the dTMT and pTMT were poorly correlated and marginally nonsignificant after applying the Bonferroni correction. Possible explanations for this occurrence include the inclusion of outliers in the data analysis, explained more in the limitation section, and overall score differences between the dTMT and pTMT, as the latter was administered after the dTMT, which meant that the participants could have gotten used to the testing conditions which could have an influence over performance and scoring. Additionally, since the HC participants were given both the paper and digital versions of the TMT, there could have been the influence of practice and/or learning effects, as both versions are very similar. Thus, it should be considered the HC group quickly adapted to the pTMT compared to their first chance with the dTMT resulting in the HCs taking less time to complete the pTMT. This could also affect the total time scores, however, the TMT A and TMT B both assess distinctive cognitive domains, with the TMT B

being more demanding, thus, participants took longer time to complete part B's of both the dTMT and pTMT, which would also affect the outcome measure. Moreover, the TMT B assesses parts of executive functioning and the ratio score mainly measures only domains of executive functioning, thus, considering that the TMT A has no influence from any cognitive abilities of executive functioning, the ratio score was probably affected (Specka et al., 2021; Arbuthnott & Frank, 2000; Corrigan & Hinkeldey, 1987; Llinàs-Reglà et al., 2016; Sánchez-Cubillo et al., 2009). This phenomenon, in addition to the practice effects, could have had an effect on the poor correlation.

However, even with high correlations between the completion times of the dTMT and pTMT, the mean values for the dTMT for all three pairs were comparatively higher. This suggests that on average, participants took more time in seconds to complete the digital version of the neuropsychological test, which has been similarly noted in other research (Park & Schott, 2021). This could be the result of participants taking time to adjust to the use of the stylus and tablet, especially when a large portion of the participants were belonging to an older age category who were taking the test for the first time. As Park & Schott (2021) stated in their research, older participants, aged between 50 and 82, have less comfort and familiarity with technology unlike their younger counterparts, which is a plausible explanation for the present study as well since majority of the HC group comprised of individuals being 40+ years of age, and the average age being 55. Conclusively, it can be claimed that this particular dTMT does, indeed, have comparable traits with the traditional pTMT, as the highly correlated completion times of both tests attest to this notion. Due to the practice effects, the correlations found could be lower than is the true case, as there could be an underestimation of the effect. Considering that scoring is heavily dependent on how long it takes to complete the TMT, it was slightly discouraging to see the dTMT performance to take longer to complete than the pTMT, since this would imply that the traditional psychometric

data would not be applicable to the dTMT of this current research. Nevertheless, with the utilization of the DiaNAH Test Battery, data is being quickly collected and processed in order to create comparable psychometric data for the dTMT. Moreover, with only slight differences between the dTMT and pTMT, the present study was able to further extrapolate new outcome measures, the intraindividual variability scores, in order to stretch the diagnostic potential of the TMT along with the traditional outcome measures.

RQ 2: Do the IIV Measures Provide Added Value to the digital version of the TMT?

This study's second research question focused on whether the IIV measures calculated from the dTMT scores provided any additional diagnostic or clinical value in addition to what is already gleaned from the traditional scores of the TMT. There were strong relationships between the IIV measures MD A, ISD A, and rISD A and the dTMT A, which may indicate that the longer a participant took to complete the dTMT A, the higher their score for their respective IIV measure. Longer time taken to complete the TMT would result in higher TMT scores and would signify a deficiency from normal cognitive functioning (Llinàs-Reglà et al., 2016). Since there is a positive association between IIV scores and dTMT completion time, it can also be suggested that a higher completion time would result in a higher IIV score, and a higher IIV score would denote an increased risk in cognitive dysfunction. This would be an appropriate conclusion considering Burton et al., (2006) were able to provide evidence that there is an association between an increase in neurological dysfunction and intraindividual variability. Furthermore, there were moderate relationships between the MD A, ISD A, and rISD A and dTMT B which is in line of expectation as the IIV measures were calculated from the dTMT A. Thus, considering the differences in cognitive assessments between parts A and B of the TMT, one would expect the IIV measures of part A to be slightly less correlated with the dTMT B and vice versa. Indeed, the IIV measures of part B had strong relationships with the dTMT B compared to the moderate

correlation of the dTMT A with the dTMT B. This is a good indication that indeed the IIV measures may provide some added potential to the cognitive assessment of an individual, considering none of the correlations equaled to 1, so it can be assumed the IIV measures do provide additional information for consideration, in addition to the traditional measures. Accordingly, some studies have shown that intraindividual variability in testing has potential to reflect neuropsychological and performance decline (Hultsch et al., 2000; Bleiberg et al., 1997; Collins & Long, 1996; Stuss et al., 1989, 1994). Thus, there is potential for IIV measures calculated from the present study's dTMT to account for some shift in mental functioning, which was explored in the third research question. Interestingly, however, the ratio score of the dTMT had very low to negative and non-significant relationships with all the part A's IIV measures while having moderate, positive relationships with all of the part B's IIV measures. A possible explanation could be found in the study of Sanchez-Cubillo et al. (2009). These authors conducted research to assess the underlying cognitive mechanisms behind the TMT A, TMT B, and its derived scores. According to the study, much of the TMT A's variance was explained by perceptual speed and visual search while task switching and working memory accounted for TMT B's variance, both of which did not correlate with the ratio score derived from the TMT. Research suggests that the ratio score is the strongest measure of executive functioning as it is derived from working memory and inhibition control, without the influence of processing speed (Specka et al., 2021; Arbuthnott & Frank, 2000; Corrigan & Hinkeldey, 1987; Llinàs-Reglà et al., 2016; Sánchez-Cubillo et al., 2009). This explanation seems applicable to the results of the present study as well, since the dTMT B assesses higher executive functioning than the dTMT A, the ratio score probably diluted the effects of part A, which makes it less correlated with part A than with part B. Therefore, a low correlation was expected between the IIV measures and the dTMT A, and moderate correlations were assumed between the IIV scores and the dTMT B. All in all, the results

provided insight for the current research to be able to believe that the IIV measures are able to provide essential and novel information along with the traditional scores calculated from the TMT. It also laid the foundation for the IIV scores' capability to potentially differentiate between healthy populations from those with ABIs which was further appraised in the third research question.

RQ 3: Do the ABI group show greater IIV than the HC group on the TMT IIV measures?

The third research question in the present study explored whether the IIV measures could discriminate between healthy controls and those with acquired brain injury by reflecting those with ABIs having higher scores than the HCs. This was conducted through two separate Independent Samples T-Tests; one with the matched participants sample as the independent group and the MD and ISD variables as the dependent group and another with all the participants as the independent group and just the rISD variables as the dependent group. For the MD and ISD variables, the study was able to use a cohort of 220 participants with and without ABI who were matched one-on-one with those with similar educational, age, and gender backgrounds in order to control for the confounding effects of those three demographics. Research suggests that age does have a confounding effect on performance on the TMT, especially for the TMT B (Woods et al., 2015; Zalonis et al., 2008), with performance diminishing as age gets higher. Education level also seems to have an effect on TMT performance as performance may increase or decrease depending on intelligence (Tombaugh, 2004; Spreen & Strauss, 1998). Overall, demographic data such as age, gender, and education level seem to have an effect on TMT performance, and the present study wanted to assess the viability of the IIV measures without the influence of confounding factors in the dTMT. The results of the current research show that for each of the MD and ISD variables, the ABI patient group had significantly higher scores, on average, than the

healthy control group. This could suggest that the present study's IIV measures are able to discriminate between ABI and HC groups, with a higher score indicating compromised neurological functioning. Likewise, other studies corroborate the aforementioned finding, as certain neurological conditions, such as traumatic brain injury, have higher IIV scores than the healthy cohort (Stuss et al., 2003; Bielak & Anstey, 2019). Moreover, research has documented that baseline IIV scores have a potential to predict future neuropsychological deterioration, such as mild cognitive impairment (Bielak et al., 2010; Cherbuin et al., 2010; Bielak & Anstey, 2019) and dementia (Kochan et al., 2016; Tales et al., 2012; Bielak & Anstey, 2019). For the rISD variable, there was no matched group as the variable is standardized in itself; the residualized ISD scores were converted with normalized scores within a standard population in order to maintain uniformity through the removal of within and in between person performance differences through the use of the residuals (Ali et al., 2019). Thus, the whole sample population was used for this variable as there are no large confounding effects due to the standardization applied. Results clearly indicated that the ABI group, again, scored higher than the HC group, for both rISD A and rISD B, which means that all of this study's IIV variables are suggested to differentiate between brain injured persons and healthy participants. Indeed, a study conducted by Bielak & Anstey (2019) documented that those with a higher rISD score, on average, showed greater decline in functioning over time. The rISD variables had large effect sizes as well, which means that the IIV measure showed a great difference between the ABI and HC participants. For the purposes of the present study, seeing such differences between healthy participants and those with ABI gives a strong indication that all three of the IIV measures were able to separately identify the two groups, and a higher score in any of the IIV variables indicated brain injury. By establishing the differential capabilities of the IIV measures, the current research is able to support its ability to distinguish between ABI and non-ABI participants, however, there could

be some predictor variables that have influence over the measures in its aptitude to differentiate. Thus, the last research question of the present study assessed potential predictors that might have had, if any, effect on the IIVs.

RQ 4: What predictor variables influence IIV in ABI and HC participants?

The final research question studied was to see whether certain predictor variables had an impact on the IIV variables. This was conducted by separate multiple regression models for the ABI and HC groups. For the healthy participants, age, gender, handedness, and education level were analyzed in the model. Currently, there are inconsistencies on the effects of gender on the performance on the TMT (Woods et al., 2015) thus the present study would like to assess whether these inconsistencies are also apparent in the dTMT. While age and a higher educational level had significant relationships for all of the dTMT B's IIV measures, age was the only variable that had a significant relationship for all of the dTMT A's IIV measures. Interestingly, the rISD A and B measures had no significant relationships with handedness. There are some conclusions that can be possibly drawn from the influence of predictor variables in healthy control groups. Firstly, the higher the age of the participant, the higher the score of the IIV measures. This finding is consistent with other research which have shown that IIV scores tend to increase with age even in healthy cohorts (Bielak & Anstey, 2019; Hultsch et al., 2002; MacDonald, Hultsch, & Dixon, 2003). Schretlen et al. (2003) found that the MD IIV measure also increased in value as age increased even in healthy populations. Naturally, cognitive performance does decline with age in healthy populations as seen in a longitudinal study by Bielak & Anstey (2019). The researchers reported that the association of higher IIV scores and lower cognitive functioning was stronger in their young cohort compared to their middle and older cohorts as cognitive decline was more apparent, from individual baseline, in the younger sample in a span of over 8 years. Thus, this shows that it is naturally expected to decline in cognitive performance as

one gets older, regardless of the absence of an ABI. Hence, the results of the current research are in accordance to other studies which reflect age as an influential predictor variable for IIV measures. This could be explained by the idea that an IIV measure is seen as a developmental phenomenon (Bielak & Anstey, 2019) and can assume longitudinal cognitive changes (Bielak & Anstey, 2019; Kochan et al., 2016) which is apparent as well in older, healthy populations because of natural cognitive decline as one gets older. Secondly, since the TMT B tests higher executive functioning than the TMT A, it is insinuated that a higher education level would decrease as the IIV value increases, since education level has been a constant predictor for performance on the TMT (Tombaugh, 2004; Woods et al., 2015). This current research's result also corroborates the aforementioned phenomenon as there was an indication that there is some relation between the dTMT B's IIV and education level; specifically, a higher education level. Tombaugh's (2004) study also concluded that there were only effects of education and age, but not gender, on the performance on the TMT, which seems to be the case in the present study as well. For the handedness variable, there was no indication of influence on the performance on the dTMT, which is contradictory to the study conducted by Bracken et al. (2018) as they found effects of both left and right handedness on their iPad version of the dTMT and the pTMT in a sample of currently enrolled students starting at the age of 18. This discrepancy could be due to their cohort consisting of a smaller sample size than the present study as well as the fact that their cohort consisted of college students who are well versed in test taking and technology in comparison to this current study's older, mixed cohort.

Subsequently, the influence of age, gender, handedness, education level, ABI etiology, the left, right, or both sides of the brain affected by the ABI, time since the last ABI, and if the binocular visual field was affected for each IIV measure was analyzed within the ABI population. In the ABI cohort, the significant predictors were variant for the IIV

measures, with age being the only constant predictor for all of the IIVs, except for the rISD measures, as age was not a predictor that was tested with it. For the MD B and ISD B measures, gender and ABI etiology were significant, in addition to age. In the ISD A model, binocular visual field was significant alongside age. The rISD B model had ABI etiology as the only significant predictor. As mentioned previously, age being a significant predictor for all the IIVs is in accordance with previous research regarding higher age and subsequent diminishing functioning. Interestingly enough, education level was no longer a predictor variable for any of the IIVs for the ABI cohort. This could be in part that the ABI cohort were older on average and naturally undergo cognitive decline and their ABI diagnosis probably also affected their performance. Thus, the effects of education might have diminished due to the aforementioned confounding factors. Moreover, most of the ABI participants had a low or middle education level, which affects performance. Indeed, a study conducted by Tombaugh (2004) reflected that as age increased and education level decreased, participants performed worse on the TMT. It is thought-provoking to see ABI etiology only influencing the MD B, ISD B, and rISD B IIV measures, especially those with hemorrhagic cardiovascular accident, and a possible explanation could be that the neurological impairment in the 17 participants with hCVA coincided with the cognitive domains affected in the performance of the dTMT-B, which was screened in the subsequent IIV measures. Another possible explanation could be that the effects a hCVA would have in everyday functioning would be variant from person to person, as brain injury in one area would not cognitively or physically affect someone else with a hCVA in the same way. Indeed, studies state that after an event of a stroke, various cognitive domains are affected depending on stroke etiology (Torrisi et al., 2019; Zucchella et al., 2014). So, since it is not known what neurological functioning has been affected by the brain injury, it cannot be assumed how every ABI participant would be affected in their TMT performance, especially when various domains are assessed while performing the TMT.

Other research suggest that processing speed and executive functioning are most commonly affected post-stroke (Kliem et al., 2021; Cumming et al., 2012). To the best of our knowledge, there is little research on different CVA types and IIV measures derived from a dTMT. Therefore, future research should delve deeper into these IIV measures and their capacity in differentiating between other disorders, etiologies, and diseases.

Limitations

The results of this current study should be approached with some caution as the conscious decision to include outliers in the data analysis was made. This is because the removal of any outlier would have purposefully excluded participants with extreme scores in the dTMT or pTMT, which would then tailor the results to reflect statistical significance, and remove legitimate observations. This would then become an erroneous decision as the study would no longer be representing a real-world population analysis. Thus, in order to avoid data manipulation, bias, and Type I Errors (Gress et al., 2018), the data was treated with logarithm transformations in order to meet the assumptions for the multiple regression analyses. However, the 'Time before onset' predictor variable remained independently nonlinear for each IIV measure. Nonetheless, due to the residual plots for the model being normal and linear, this issue was overlooked. Another limitation includes the purchase and use of specific hardware, the Wacom 24-inch screen tablet from Metrisquare B.V., and DiagnoseIS software.

Moreover, the DiaNAH Test Battery that includes the dTMT from the current research has to be purchased separately as well. The aforementioned limitations could be very costly to purchase and difficult to justify the cost effectiveness to your employer or organization. Also, the constant need to update the software and test battery versions is not only cost inefficient, but also time consuming to learn and adapt to the new formatting as an individual practitioner or a whole clinic. Additionally, normative and psychometric data is currently being collected, so as of the present study, there are no standardized information

available for comparison, however, normative data for the Dutch population is being updated and will be soon available for use and include the benefit of automatic test scoring. A third limitation includes the possibility of practice and/or learning effects within the healthy participants as they took both TMT versions subsequently. Thus, there is a chance that some participants were able to perform better in the pTMT as they may have gotten used to the test through their first trial of the dTMT which could have influence over the size of the effect found in the correlational analysis between the two. Namely, the participants who took the pTMT could have finished the test quicker than their attempt at the dTMT, which would affect their scoring, as scoring relies on time. However, an overestimation of results would be preferred over an underestimation in order to avoid undetected neurological decline.

Interestingly, Bracken et al. (2019) found that the spatial dimensions of their screen versus the paper TMT could have had an effect on performance. This could hold true in this research as well, considering the screen's dimensions were larger than the paper-based test, which would affect the distance between dispersed stimuli. Lastly, while some may argue the irrelevance of the Bonferroni adjustment in clinical use (Perneger, 1998), as the argument is that the Bonferroni method relies on the generic null hypothesis, which is the "wrong hypothesis" as it is of no use to the researcher. Perneger argues that the risk of type II errors is increased through the method, and that important results are then seen as non-significant and the statistical dependency remains on the number of other tests performed. Nonetheless, the present study applied the Bonferroni method in order to avoid the possibility of rejecting a true null hypothesis, as it is the safer statistical mistake to occur in comparison to a type II error.

Future Direction

With the novelty of the current research's dTMT, the future direction for further studies has many possibilities. Firstly, the transferable qualities from the pTMT to the dTMT

should be replicated in other research in order to establish corroboration and psychometric data, such as norms, validity and reliability. The first indications of psychometric data for the dTMT from the DiaNAH Test Battery (de Vries et al., 2017) are currently being collected within the Dutch population, and would be available for use in the near future, allowing for the added benefits of automatic scoring and reporting. Nevertheless, there have been some studies which digitalized the TMT with other methods, such as an android application-based test on a 12-inch multi-touch screen Samsung Galaxy Note Pro (Park & Schott, 2021) or a 9.7-inch iPad with a stylus (Bracken et al., 2019), which provides some psychometric insight. The iPad research conducted by Bracken et al. (2019) showed that Part A of their dTMT did not provide efficient test-retest reliability, however, Part B showed promising test-retest reliability in most of their groups. On the other hand, their Part B did not result in acceptable concurrent validity while Part A did. However, the authors of the study did point that the promising validity in Part A had no significance without acceptable reliability. Conversely, the research conducted by Park & Schott (2021) resulted in adequate reliability, validity, and high sensitivity. Clearly, further research is necessary in order to figure out the psychometric information of computerized TMT testing, including the DiaNAH Test Battery (de Vries et al., 2017), in order to pertinently assess whether or not digitalized testing could be used as a replacement for traditional neuropsychological testing.

Next, intraindividual variability as calculated in the present study should be replicated in future research using this current dTMT, in order to establish further evidence for the use of the three IIV measures in the context of predicting neurological functioning. It would also be interesting to see whether this dTMT would be able to distinguish between malingerers and genuine performance like in other versions of the dTMT (Woods et al., 2015). The current study was able to indicate that these IIV measures have the possibility of distinguishing between participants with ABI and those in the HC group, which should be

further replicated and tested within other disorders to gauge whether they can differentiate between the dysfunctional and healthy, and what higher or lower scores actually translate to in real world functioning through longitudinal analyses of the sample population (Bielak & Anstey, 2019). Furthermore, with regards to the three different IIV measures, there are some considerations to review, even though all three were able to differentiate between the ABI and HC cohorts, because there are clear differences amongst the IIV measures in theory and practicality. For example, the ISD measure does not control for mean reaction time, and since the ISD is calculated the standard deviation (SD) over ISIs, the value of the IIV can be skewed (Bielak & Anstey, 2019; Hultsch et al., 2008). This is because the SD is strongly correlated with the mean, so a participant scoring high or low could be because of the correlational effect between the SD and mean. Thus, the ISD measure would not be reflective of true variability (Bielak & Anstey, 2019; Hultsch et al., 2008). Also, the method of calculating the ISD is questionable in itself, as the value would be reliant on the main effect of the ISD, effect of the mean performance, and the interaction of the two, which means the ISD score could be influenced by the aforementioned effects, and not just variability (Bielak & Anstey, 2019; Hultsch et al., 2008). Therefore, it would be recommended to follow the calculation of the rISD measure, inspired in part by Hultsch et al (2000; 2008), as this is the residualized version of the raw score. The rISD is more apt in apportioning the influence of the interaction and main effects of other factors, making the variability score more accurate. With regards to the MD measure, the only limitation with it is that those who show pronounced IIV at the uppermost distributed scoring on the TMT might be less likely to reflect their IIV score since mean scores are already "close to their ceilings (Schretlen et al., 2002)." This could be explained by the fact that some neuropsychological tests measure a person's ability rather than deficit, in which the MD fails to establish in those people who score really high. Hence, the choice to use which IIV measure relies on what the researcher

wants, which neuropsychological test is being used, and whichever measure suits their study best. Overall, however, the rISD seems to be the most flexible in analysis, while the MD depends on the testing used or to interpret the results with the hiccup in mind. With the increasing use of computerized testing and technology in neuropsychology, the present study's dTMT has a lot of potential that needs to be further replicated and researched. The application of the dTMT and its IIV measures in clinical practice would revolutionize the diagnostic and rehabilitation competencies of neuropsychology. People could be informed about their cognitive health longitudinally and have it profiled over time to note progress and decline of disorder or disease. Treatment can initiate at earlier times which could preserve disease progression or compensatory skills can be taught in advance in order to be prepared for life long changes. All of these benefits could preserve and increase the life expectancy and satisfaction of patients, which would now can be possible through the extended research of the dTMT and intraindividual variability.

Conclusion

The research conducted in this study was able to extend the possibilities of traditional paper and pencil testing, opening doors to the many potential opportunities offered through the digitalization of traditional testing, for prognosis, diagnostic, and rehabilitation purposes. The dTMT garnered from the DiaNAH Test Battery has been shown to be comparable and consistent with the pTMT, but with more potential offered. The dTMT was valuable for clinical use in addition to the traditional outcome measures, such as the calculation of the intraindividual variability of a participant. The digitalization and assessment of the IIV was conducted through four research questions analyzed in the present study. The first research question found the dTMT to have variant completion times to the pTMT, which laid the foundation for the present study to further assess why there were noticeable differences and if there was any added value in the dTMT. Thus, the second research question utilized a

correlational analysis to extrapolate the newfound potential of the dTMT in addition to the pTMT and when that was noticed, the correlations helped in order to justify the calculation of IIV measures from the dTMT. Subsequently, the third research question wanted to gauge whether the IIVs had the capability to differentiate between those with and without an ABI, which was confirmed in our results as the ABI cohort scored higher IIV values than the HC. The fourth research question found 'Age' to be a predictor variable for the MD and ISD measures in both participant groups. Education level had a predictor effect for the HC group for all the IIVs, excluding the rISD while stroke etiology had an effect on the dTMT B for the IIV measures.

The current research was the initial one to establish validation of the dTMT and its potential through the IIV outcome measures. Other immediate benefits include decrease in administration time and effort, easy storage of information, standardization of the whole assessment process, and efficient scoring and patient profiling, all unattainable through the pTMT. Moreover, the IIV measures calculated from the dTMT seems to have potential to differentiate between normal and brain injured functioning, however, there is a lot of research needed in order to understand what the distinction between the ABI and HC mean, how it translates into real world functioning, and in what capacity can it be used in diagnostics. Nevertheless, the validation and establishment of one of the first research into the dTMT and intraindividual variability calculated in three different ways is groundbreaking in neuropsychological care. Through the continued development of the findings of this dTMT and IIV, there is no doubt the added potential would blossom and be essential in future care. Consequently, the present study has cemented some of the foundation into unlocking the many intraindividual, diagnostic possibilities that can be gleaned from the TMT.

References

Ali, S., Macoun, S. J., Bedir, B., & MacDonald, S. W. (2019). Intraindividual variability in children is related to informant ratings of attention and executive function. *Journal of Clinical and Experimental Neuropsychology*, *41*(7), 740–748. https://doi.org/10.1080/13803395.2019.1617249

Amodio, P., Wenin, H., Del Piccolo, F., Mapelli, D., Montagnese, S., Pellegrini, A., Musto, C., Gatta, A., & Umiltà, C. (2002). 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. *Aging Clinical and Experimental Research*, *14*(2), 117–131. https://doi.org/10.1007/bf03324425

Arbuthnott, K., & Frank, J. (2000). Trail making test, part B as a measure of executive control: Validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology*, *22*(4), 518–528. https://doi.org/10.1076/1380-3395(200008)22:4;1-0;ft518

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology, 27*(3), 362-373. doi:10.1093/arclin/acs027

Bezdicek, O., Motak, L., Axelrod, B. N., Preiss, M., Nikolai, T., Vyhnalek, M., Poreh, A., & Ruzicka, E. (2012). Czech version of the trail making test: Normative Data and clinical utility. *Archives of Clinical Neuropsychology*, *27*(8), 906–914. https://doi.org/10.1093/arclin/acs084

Bielak, A. A., & Anstey, K. J. (2019). Covariation of intraindividual variability in cognitive speed and cognitive performance across young, middle, and older adulthood. *Developmental Psychology, 55*(5), 994-1004. doi:10.1037/dev0000688

Bielak, A. A., Hultsch, D. F., Strauss, E., MacDonald, S. W., & Hunter, M. A. (2010). Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. *Neuropsychology, 24*(6), 731–741. https://doi.org/10.1037/a0019802

Bilder, R. M., & Reise, S. P. (2018). Neuropsychological tests of the future: How do we get there from here? *The Clinical Neuropsychologist*, *33*(2), 220–245. https://doi.org/10.1080/13854046.2018.1521993

Björngrim, S., van den Hurk, W., Betancort, M., Machado, A., & Lindau, M. (2019). Comparing traditional and digitized cognitive tests used in standard clinical evaluation – A study of the digital application minnemera. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02327

Bracken, M. R., Mazur-Mosiewicz, A., & Glazek, K. (2018). Trail Making Test: Comparison of paper-and-pencil and electronic versions. *Applied Neuropsychology: Adult*, *26*(6), 522–532. https://doi.org/10.1080/23279095.2018.1460371

Burton, C. L., Hultsch, D. F., Strauss, E., & Hunter, M. A. (2002). Intraindividual Variability in Physical and Emotional Functioning: Comparison of Adults With Traumatic Brain Injuries and Healthy Adults. *The Clinical Neuropsychologist*, *16*(3), 264–279. https://doi.org/10.1076/clin.16.3.264.13854

Cangoz, B., Karakoc, E., & Selekler, K. (2009). Trail making test: Normative data for Turkish elderly population by age, sex and Education. *Journal of the Neurological Sciences*, *283*(1-2), 73–78. https://doi.org/10.1016/j.jns.2009.02.313

Cavaco, S., Goncalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Trail making test: Regression-based norms for the Portuguese population. *Archives of Clinical Neuropsychology*, *28*(2), 189–198. https://doi.org/10.1093/arclin/acs115

Cernich, A., Brennana, D., Barker, L., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology, 22,* 39-48. doi:10.1016/j.acn.2006.10.004

Chapman, R. M., Mapstone, M., McCrary, J. W., Gardner, M. N., Porsteinsson, A., Sandoval, T. C., Guillily, M. D., DeGrush, E., & Reilly, L. A. (2010). Predicting conversion from mild cognitive impairment to alzheimer's disease using neuropsychological tests and multivariate methods. *Journal of Clinical and Experimental Neuropsychology, 33*(2), 187–199. https://doi.org/10.1080/13803395.2010.499356

CHAYTOR, N., SCHMITTEREDGECOMBE, M., & BURR, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*(3), 217–227. https://doi.org/10.1016/j.acn.2005.12.002

Cherbuin, N., Sachdev, P., & Anstey, K. J. (2010). Neuropsychological predictors of transition from healthy cognitive aging to mild cognitive impairment: The path through life study. *The American Journal of Geriatric Psychiatry*, *18*(8), 723–733. https://doi.org/10.1097/jgp.0b013e3181cdecf1

Cohen, J. (1988). *Statistical Power Analysis for the behavioral sciences*. L. Erlbaum Associates.

Collins, L. (1996). Visual reaction time and its relationship to neuropsychological test performance. *Archives of Clinical Neuropsychology*, *11*(7), 613–623. https://doi.org/10.1016/0887-6177(97)81255-3

Corrigan, J. D., & Hinkeldey, N. S. (1987). Relationships between parts A and B of the trail making test. *Journal of Clinical Psychology*, *43*(4), 402–409. https://doi.org/10.1002/1097-4679(198707)43:4<402::aid-jclp2270430411>3.0.co;2-e

Cumming, T. B., Marshall, R. S., & Lazar, R. M. (2012). Stroke, cognitive deficits, and rehabilitation: Still an incomplete picture. *International Journal of Stroke*, *8*(1), 38–45. https://doi.org/10.1111/j.1747-4949.2012.00972.x

Dahmen, J., Cook, D., Fellows, R., & Schmitter-Edgecombe, M. (2017). An analysis of a digital variant of the Trail Making Test using machine learning techniques. *Technology and Health Care*, *25*(2), 251–264. https://doi.org/10.3233/thc-161274

de Haan, G. A., Melis-Dankers, B. J., Brouwer, W. H., Tucha, O., & Heutink, J. (2016). Correction: The effects of compensatory scanning training on mobility in patients with homonymous visual field defects: A randomized controlled trial. *PLOS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0165863

de Vries, S. M., Tucha, O., Melis-Dankers, B. J., Vrijling, A. C., Ribbers, S., Cornelissen, F. W., & Heutink, J. (2020). The Birthday Party Test (BPT): A new picture description test to support the assessment of simultanagnosia in patients with acquired Brain Injury. *Applied Neuropsychology: Adult*, *29*(3), 383–396. https://doi.org/10.1080/23279095.2020.1763998

Ewers, M., Walsh, C., Trojanowski, J. Q., Shaw, L. M., Petersen, R. C., Jack, C. R., Feldman, H. H., Bokde, A. L. W., Alexander, G. E., Scheltens, P., Vellas, B., Dubois, B., Weiner,

M., & Hampel, H. (2012). Prediction of conversion from mild cognitive impairment to alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging*, *33*(7). https://doi.org/10.1016/j.neurobiolaging.2010.10.019

Feigenson, J. S., McCarthy, M. L., Greenberg, S. D., & Feigenson, W. D. (1977). Factors influencing outcome and length of stay in a stroke rehabilitation unit. part 2. comparison of 318 screened and 248 unscreened patients. *Stroke*, *8*(6), 657–662. https://doi.org/10.1161/01.str.8.6.657

Fellows, R. P., Dahmen, J., Cook, D., & Schmitter-Edgecombe, M. (2016). Multicomponent analysis of a Digital Trail Making Test. *The Clinical Neuropsychologist*, *31*(1), 154–167. https://doi.org/10.1080/13854046.2016.1238510

Fine, E. M., Delis, D. C., & Holdnack, J. (2011). Normative adjustments to the D-KEFS trail making test: Corrections for Education and vocabulary level. *The Clinical Neuropsychologist*, *25*(8), 1331–1344. https://doi.org/10.1080/13854046.2011.609838

Frias, C. M., Dixon, R. A., & Camicioli, R. (2012). Neurocognitive Speed and Inconsistency in Parkinsons Disease with and without Incipient Dementia: An 18-Month Prospective Cohort Study. *Journal of the International Neuropsychological Society*, *18*(4), 764-772. doi:10.1017/s1355617712000422

Gilhotra, J. S., Mitchell, P., Healey, P. R., Cumming, R. G., & Currie, J. (2002). Homonymous visual field defects and stroke in an older population. *Stroke*, *33*(10), 2417–2420. https://doi.org/10.1161/01.str.0000037647.10414.d2

Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacona, M., & Capitani, E. (1996). Trail making test: Normative values from 287 normal adult controls. *The*

Italian Journal of Neurological Sciences, *17*(4), 305–309. https://doi.org/10.1007/bf01997792

Gomar, J. J. (2011). Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to alzheimer disease in patients in the alzheimer's disease neuroimaging initiative. *Archives of General Psychiatry*, *68*(9), 961. https://doi.org/10.1001/archgenpsychiatry.2011.96

Gorus, E., Raedt, R. D., Lambert, M., Lemper, J., & Mets, T. (2008). Reaction Times and Performance Variability in Normal Aging, Mild Cognitive Impairment, and Alzheimers Disease. *Journal of Geriatric Psychiatry and Neurology*, *21*(3), 204-218. doi:10.1177/0891988708320973

Gress, T. W., Denvir, J., & Shapiro, J. I. (2018). Effect of removing outliers on statistical inference: Implications to interpretation of experimental data in Medical Research. *Marshall Journal of Medicine*, *4*(2). https://doi.org/10.18590/mjm.2018.vol4.iss2.9

GUALTIERI, C., & JOHNSON, L. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital signs. *Archives of Clinical Neuropsychology*, *21*(7), 623–643. https://doi.org/10.1016/j.acn.2006.05.007

Halstead, W. C. (1947). Brain and Intelligence; a quantitative study of the frontal lobes. *University of Chicago Press*.

Hanks, R. A., Millis, S. R., Ricker, J. H., Giacino, J. T., Nakese-Richardson, R., Frol, A. B., Novack, T. A., Kalmar, K., Sherer, M., & Gordon, W. A. (2008). The predictive validity of a brief inpatient Neuropsychologic battery for persons with Traumatic Brain Injury. *Archives of Physical Medicine and Rehabilitation*, *89*(5), 950–957. https://doi.org/10.1016/j.apmr.2008.01.011

HASHIMOTO, R. Y. U. S. A. K. U., MEGURO, K. E. N. I. C. H. I., LEE, E. U. N. J. O. O., KASAI, M. A. R. I., ISHII, H. I. R. O. S. H. I., & YAMAGUCHI, S. A. T. O. S. H. I. (2006). Effect of age and education on the trail making test and determination of normative data for Japanese elderly people: The tajiri project. *Psychiatry and Clinical Neurosciences*, *60*(4), 422–428. https://doi.org/10.1111/j.1440-1819.2006.01526.x

Hester, R. L., Kinsella, G. J., Ong, B., & McGregor, J. (2005). Demographic influences on baseline and derived scores from the trail making test in healthy older Australian adults. *The Clinical Neuropsychologist*, *19*(1), 45–54. https://doi.org/10.1080/13854040490524137

Hultsch, D. F., Macdonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*, *14*(4), 588–598. https://doi.org/10.1037/0894-4105.14.4.588

Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(2). https://doi.org/10.1093/geronb/57.2.p101

Kane, R. L., & Kay, G. G. (1992). Computerized assessment in neuropsychology: A review of tests and test batteries. *Neuropsychology Review*, *3*(1), 1-117. doi:10.1007/bf01108787

Karimpoor, M., Churchill, N. W., Tam, F., Fischer, C. E., Schweizer, T. A., & Graham, S. J. (2017). Tablet-based functional MRI of the trail making test: Effect of Tablet

Interaction mode. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00496

Kliem, E., Gjestad, E., Ryum, T., Olsen, A., Thommessen, B., Indredavik, B., Bieliauskas, L., Due-Tønnessen, P., Fladby, T., & Grambaite, R. (2021). The relationship of psychiatric symptoms with performance-based and self-reported cognitive function after ischemic stroke. *Journal of the International Neuropsychological Society*, *28*(1), 35–47. https://doi.org/10.1017/s1355617721000187

Kochan, N. A., Bunce, D., Pont, S., Crawford, J. D., Brodaty, H., & Sachdev, P. S. (2016). Reaction time measures predict incident dementia in community-living older adults: The sydney memory and ageing study. *The American Journal of Geriatric Psychiatry*, *24*(3), 221–231. https://doi.org/10.1016/j.jagp.2015.12.005

Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The trail making Test, part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, *9*(2), 106–109. https://doi.org/10.1207/s15324826an0902_5

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology, 70*(2), 144. https://doi.org/10.4097/kjae.2017.70.2.144

Lamberty, G. J., Putnam, S. H., Chatel, D. M., & Bieliauskas, L. A. (1994). Derived Trail Making Test indices: A preliminary report. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology, 7*(3), 230–234.

Langenecker, S. A., Zubieta, J.-K., Young, E. A., Akil, H., & Nielson, K. A. (2007). A task to manipulate attentional load, set-shifting, and inhibitory control: Convergent validity and test–retest reliability of the parametric go/no-go test. *Journal of Clinical and*

Experimental Neuropsychology, *29*(8), 842–853. https://doi.org/10.1080/13803390601147611

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment*. Oxford University Press.

Lin, Z., Tam, F., Churchill, N. W., Lin, F.-H., MacIntosh, B. J., Schweizer, T. A., & Graham, S. J. (2021). Trail making test performance using a touch-sensitive tablet: Behavioral kinematics and electroencephalography. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.663463

Llinàs-Reglà, J., Vilalta-Franch, J., López-Pousa, S., Calvó-Perxas, L., Torrents Rodas, D., & Garre-Olmo, J. (2016). The trail making test. *Assessment*, *24*(2), 183–196. https://doi.org/10.1177/1073191115602552

Lunardini, F., Luperto, M., Daniele, K., Basilico, N., Damanti, S., Abbate, C., Mari, D., Cesari, M., Ferrante, S., & Borghese, N. A. (2019). Validity of Digital Trail making test and Bells Test in Elderlies. *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. https://doi.org/10.1109/bhi.2019.8834513

MacDonald, S. W., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychology and Aging*, *18*(3), 510–523. https://doi.org/10.1037/0882-7974.18.3.510

Makizako, H., Shimada, H., Park, H., Doi, T., Yoshida, D., Uemura, K., Tsutsumimoto, K., & Suzuki, T. (2012). Evaluation of multidimensional neurocognitive function using a tablet personal computer: Test-retest reliability and validity in community-dwelling older adults. *Geriatrics & Gerontology International*, *13*(4), 860–866. https://doi.org/10.1111/ggi.12014

Muñiz, J., & Bartram, D. (2007). Improving International Tests and Testing. *European Psychologist*, *12*(3), 206-219. doi:10.1027/1016-9040.12.3.206

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*(9), 1352–1375. https://doi.org/10.1080/00140130802170387

O'donnell, J. P., Macgregor, L. A., Dabrowski, J. J., Oestreicher, J. M., & Romero, J. J. (1994). Construct validity of neuropsychological tests of conceptual and attentional abilities. *Journal of Clinical Psychology*, *50*(4), 596–600. https://doi.org/10.1002/1097-4679(199407)50:4<596::aid-jclp2270500416>3.0.co;2-s

Onoda, Hamano, Nabika, Aoyama, Takayoshi, Nakagawa, Ishihara, Mitaki, S., Yamaguchi, Oguro, Shiwaku, & Yamaguchi. (2013). Validation of a new mass screening tool for cognitive impairment: Cognitive assessment for dementia, iPad version. *Clinical Interventions in Aging*, 353. https://doi.org/10.2147/cia.s42342

Park, S.-Y., & Schott, N. (2021). The trail-making-test: Comparison between paper-and-pencil and computerized versions in young and healthy older adults. *Applied Neuropsychology: Adult*, 1–13. https://doi.org/10.1080/23279095.2020.1864374

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of Technology in Neuropsychological Assessment. *The Clinical Neuropsychologist, 27*(8), 1328-1361. doi:10.1080/13854046.2013.834971

PERIANEZ, J., RIOSLAGO, M., RODRIGUEZSANCHEZ, J., ADROVERROIG, D., SANCHEZCUBILLO, I., CRESPOFACORRO, B., QUEMADA, J., & BARCELO, F. (2007). Trail making test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and Normative Data. *Archives of Clinical Neuropsychology*, *22*(4), 433–447. https://doi.org/10.1016/j.acn.2007.01.022

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, *316*(7139), 1236–1238. https://doi.org/10.1136/bmj.316.7139.1236

Rabin, L. A., Spadaccini, A. T., Brodale, D. L., Grant, K. S., Elbulok-Charcape, M. M., & Barr, W. B. (2014). Utilization rates of computerized tests and test batteries among clinical neuropsychologists in the United States and Canada. *Professional Psychology: Research and Practice, 45*(5), 368-377. doi:10.1037/a0037987

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology*, *19*(5), 393–394. https://doi.org/10.1037/h0044509

Reitan, R. M. (1971). Trail making test results for normal and brain-damaged children. *Perceptual and Motor Skills*, *33*(2), 575–581. https://doi.org/10.2466/pms.1971.33.2.575

Reitan, R. M., & Wolfson, D. (1986). 3/the halstead-reitan neuropsychological test battery and aging. *Clinical Gerontologist*, *5*(1-2), 39–61. https://doi.org/10.1300/j018v05n01_03

Ríos, M., Periáñez, J. A., & Muñoz-Céspedes, J. M. (2004). Attentional control and slowness of information processing after severe traumatic brain injury. *Brain Injury*, *18*(3), 257–272. https://doi.org/10.1080/02699050310001617442

Salthouse, T. A., & Fristoe, N. M. (1995). Process analysis of adult age effects on a computer-administered trail making test. *Neuropsychology*, *9*(4), 518–528. https://doi.org/10.1037/0894-4105.9.4.518

SÁNCHEZ-CUBILLO, I., PERIÁÑEZ, J. A., ADROVER-ROIG, D., RODRÍGUEZ-SÁNCHEZ, J. M., RÍOS-LAGO, M., TIRAPU, J., & BARCELÓ, F. (2009). Construct validity of the trail making test: Role of Task-switching, working memory,

inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, *15*(3), 438–450. https://doi.org/10.1017/s1355617709090626

Schober, P. A., Boer, C. A., & Schwarte, L. A. (2018). Correlation Coefficients. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ane.0000000000002864

Schretlen, D. J., Munro, C. A., Anthony, J. C., & Pearlson, G. D. (2003). Examining the range of normal intraindividual variability in neuropsychological test performance. *Journal of the International Neuropsychological Society*, *9*(6), 864–870. https://doi.org/10.1017/s1355617703960061

Seo, E. H., Lee, D. Y., Kim, K. W., Lee, J. H., Jhoo, J. H., Youn, J. C., Choo, I. L. H., Ha, J., & Woo, J. I. (2006). A normative study of the trail making test in Korean elders. *International Journal of Geriatric Psychiatry*, *21*(9), 844–852. https://doi.org/10.1002/gps.1570

Specka, M., Weimar, C., Stang, A., Jöckel, K.-H., Scherbaum, N., Hoffmann, S. S., Kowall, B., & Jokisch, M. (2021). Trail making test normative data for the German older population. *Archives of Clinical Neuropsychology*, *37*(1), 186–198. https://doi.org/10.1093/arclin/acab027

Spikman, J. M., Kiers, H. A. L., Deelman, B. G., & van Zomeren, A. H. (2001). Construct validity of concepts of attention in healthy controls and patients with chi. *Brain and Cognition*, *47*(3), 446–460. https://doi.org/10.1006/brcg.2001.1320

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.

Spreij, L. A., Gosselt, I. K., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2020). Digital neuropsychological assessment: Feasibility and applicability in patients with acquired brain injury. *Journal of Clinical and Experimental Neuropsychology*, *42*(8), 781–793. https://doi.org/10.1080/13803395.2020.1808595

Stawski, R. S., Macdonald, S. W. S., Brewster, P. W. H., Munoz, E., Cerino, E. S., & Halliday, D. W. R. (2017). A Comprehensive Comparison of Quantifications of Intraindividual Variability in Response Times: A Measurement Burst Approach. *The Journals of Gerontology: Series B*, *74*(3), 397–408. https://doi.org/10.1093/geronb/gbx115

Strauss, E., S., S. E. M., Spreen, O., & Spreen, O. (2006). *A compendium of neuropsychological tests administration, norms, and commentary*. Oxford University Press.

Stuss, D. T. (2003). Staying on the job: The frontal lobes control individual performance variability. *Brain*, *126*(11), 2363–2380. https://doi.org/10.1093/brain/awg237

Stuss, D. T., Bisschop, S. M., Alexander, M. P., Levine, B., Katz, D., & Izukawa, D. (2001). The trail making test: A study in focal lesion patients. *Psychological Assessment*, *13*(2), 230–239. https://doi.org/10.1037/1040-3590.13.2.230

Stuss, D. T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests. *Neuropsychology*, *8*(3), 316–324. https://doi.org/10.1037/0894-4105.8.3.316

Stuss, D. T., Stethem, L. L., Hugenholtz, H., Picton, T., Pivik, J., & Richard, M. T. (1989). Reaction time after head injury: Fatigue, divided and focused attention, and consistency

of performance. *Journal of Neurology, Neurosurgery & Psychiatry*, *52*(6), 742–748. https://doi.org/10.1136/jnnp.52.6.742

Tales, A., Leonards, U., Bompas, A., Snowden, R. J., Philips, M., Porter, G., Haworth, J., Wilcock, G., & Bayer, A. (2012). Intra-individual reaction time variability in amnestic mild cognitive impairment: A precursor to dementia? *Journal of Alzheimer's Disease*, *32*(2), 457–466. https://doi.org/10.3233/jad-2012-120505

Tombaugh, T. (2004). Trail making test A and B: Normative data stratified by age and Education. *Archives of Clinical Neuropsychology*, *19*(2), 203–214. https://doi.org/10.1016/s0887-6177(03)00039-8

Torrisi, M., Maresca, G., De Cola, M. C., Cannavò, A., Sciarrone, F., Silvestri, G., Bramanti, A., De Luca, R., & Calabrò, R. S. (2019). Using telerehabilitation to improve cognitive function in post-stroke survivors: Is this the time for the continuity of care? *International Journal of Rehabilitation Research*, *42*(4), 344–351. https://doi.org/10.1097/mrr.0000000000000369

Vries, S. M. D., Heutink, J., Melis-Dankers, B. J. M., Vrijling, A. C. L., Cornelissen, F. W., & Tucha, O. (2017). Screening of visual perceptual disorders following acquired brain injury: A Delphi study. *Applied Neuropsychology: Adult*, *25*(3), 197–209. https://doi.org/10.1080/23279095.2016.1275636

Wecker, N. S., Kramer, J. H., Wisniewski, A., Delis, D. C., & Kaplan, E. (2000). Age effects on executive ability. *Neuropsychology*, *14*(3), 409–414. https://doi.org/10.1037/0894-4105.14.3.409

Wei, W., Zhào, H., Liu, Y., & Huang, Y. (2019). Traditional Trail Making Test Modified into Brand-new Assessment Tools: Digital and Walking Trail Making Test. *Journal of Visualized Experiments*, (153). https://doi.org/10.3791/60456

Wiberg, B., Kilander, L., Sundström, J., Byberg, L., & Lind, L. (2012). The relationship between executive dysfunction and post-stroke mortality: A population-based Cohort Study. *BMJ Open, 2*(3). https://doi.org/10.1136/bmjopen-2011-000458

Wiberg, B., Lind, L., Kilander, L., Zethelius, B., Sundelof, J. E., & Sundstrom, J. (2010). Cognitive function and risk of stroke in elderly men. *Neurology, 74*(5), 379–385. https://doi.org/10.1212/wnl.0b013e3181ccc516

Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2015). The effects of aging, malingering, and traumatic brain injury on Computerized Trail-making Test Performance. *PLOS ONE, 10*(6). https://doi.org/10.1371/journal.pone.0124345

Zalonis, I., Kararizou, E., Triantafyllou, N. I., Kapaki, E., Papageorgiou, S., Sgouropoulos, P., & Vassilopoulos, D. (2008). A normative study of the trail making test A and B in Greek adults. *The Clinical Neuropsychologist*, *22*(5), 842–850. https://doi.org/10.1080/13854040701629301

Zucchella, C. (2014). Assessing and restoring cognitive functions early after stroke. *Functional Neurology*. https://doi.org/10.11138/fneur/2014.29.4.255

Appendix A

Assumption Section's Tables and Figures

The tables and figures below belong to the Assumptions section of the four research questions analyzed. All figures are labeled from left to right.

RQ 1: Paired Sample T-Tests to Assess Differences Between the dTMT and pTMT

Table A1 depicts the normality assumption for the Paired Sample T-Tests of the first research question. As seen in the table, all three pairs violate the normality requirement as they all fail to reject the null hypothesis.

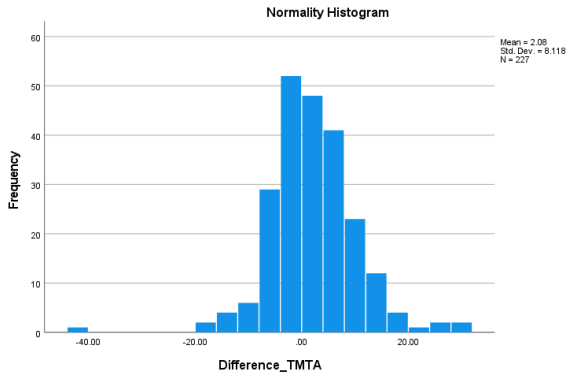
Table A1

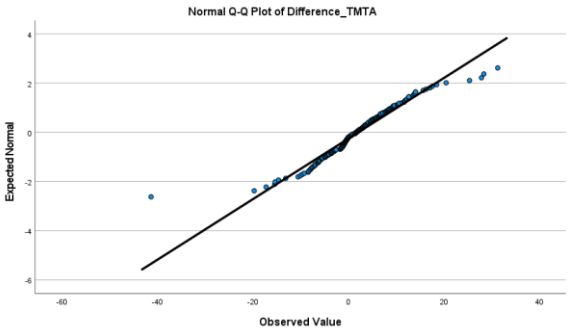
Kolmogorov-Smirnov Test of Normality Between the Differences of the dTMT and pTMT

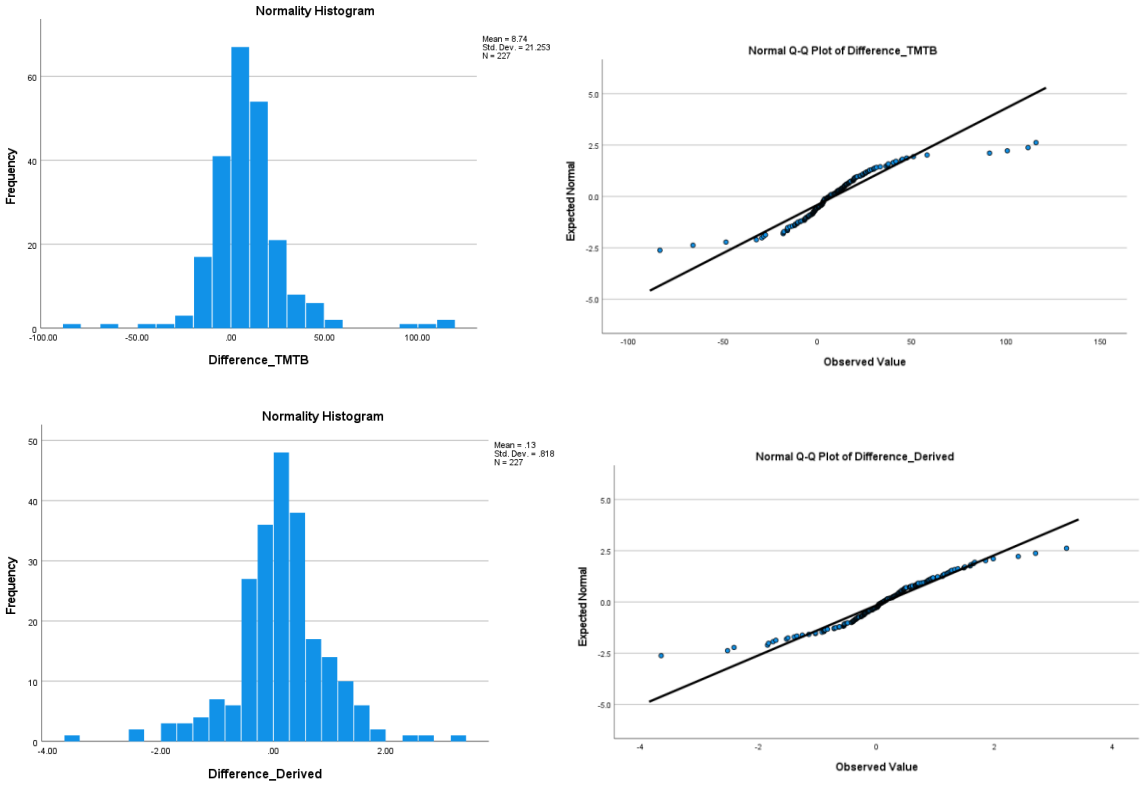
	Statistic	*df*	*p*
Difference TMT A	0.08	227	0.002
Difference TMT B	0.12	227	<0.001
Difference B/A	0.10	227	<0.001

Note. Difference TMT A refers to the subtraction of the total time of the pTMT A from the dTMT A. Difference TMT B is the difference between the total time of the dTMT B and the pTMT B. Similarly, Difference B/A referred to the difference between the dTMT's ratio score and the pTMT's ratio score. Lilliefors Significance Correction.

Figures A1 to A6, below, are corresponding histograms and scatterplots which also corroborate the normality violation.







Figures A1 to A6. Figures A1, A3, and A5 depict the difference in total time it took participants to complete part A of the pTMT and dTMT. This difference was calculated by the subtraction of the pTMT A from the dTMT A. The x-axis represents the difference score of each HC participant in seconds, while the y-axis represents the number of participants who completed in the allocated time in seconds. Figures A2, A4, and A6 depict the nonnormality of the paired sample variables due to the inclusion of outliers. Scatterplots that depict the difference of the pTMT A from the dTMT A, pTMT B from the dTMT B, and difference of the ratio scores of the pTMT from the dTMT, respectively, as a plotted value for each HC participant. As seen, the outliers detrend the plot from a linear form into a nonlinear trend.

RQ 2: Correlation Between IIV Measures and dTMT Variables

The next set of tables and figures belong to the second research question exploring the correlations between all the variables of the dTMT.

Table A2

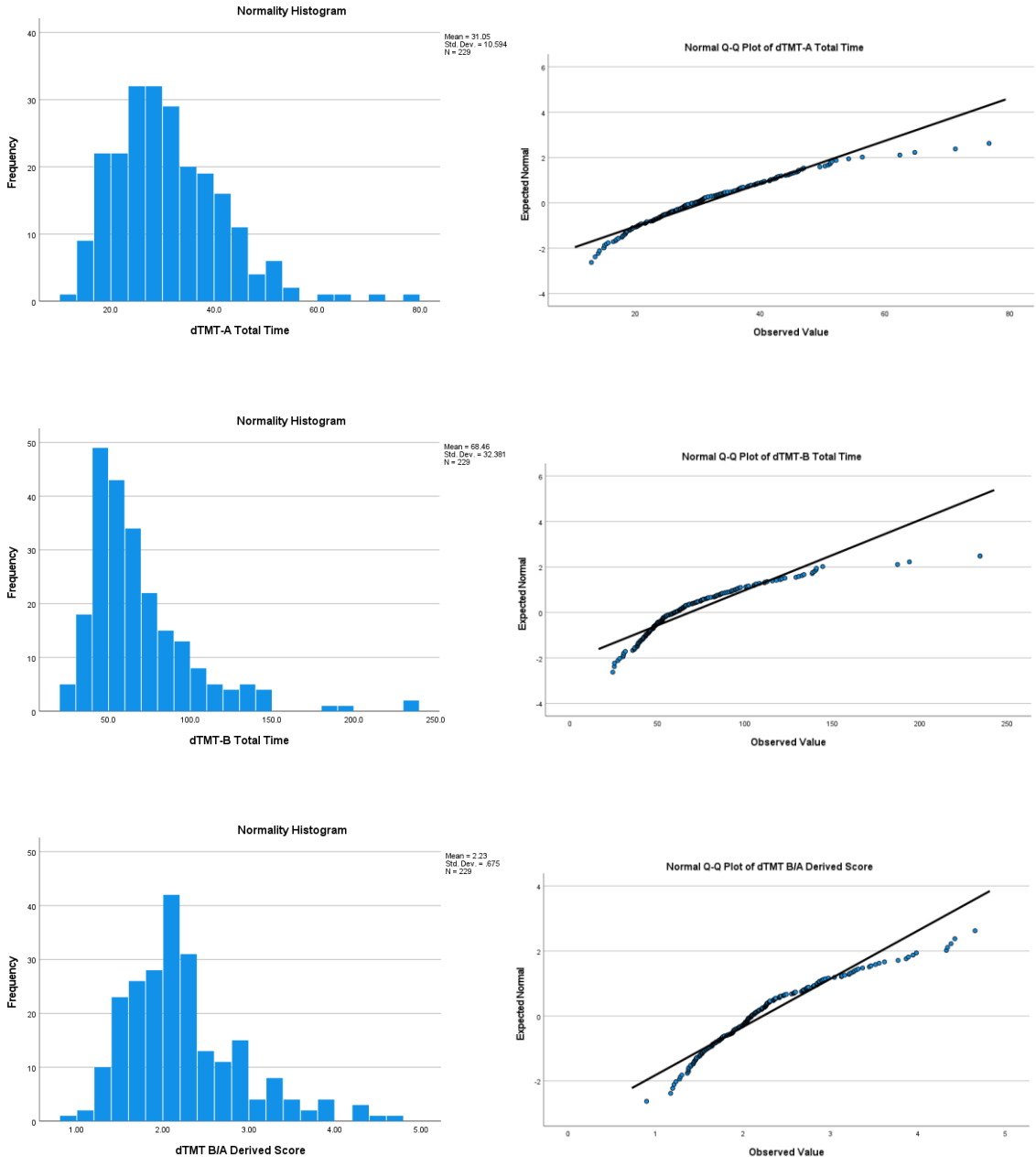
Kolmogorov-Smirnov Test of Normality Between Variables of the dTMT and IIV Measures

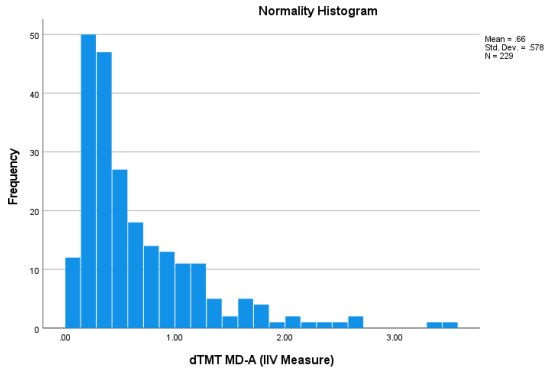
	Statistic	*df*	*p*
dTMT A	0.09	229	<0.001
dTMT B	0.15	229	<0.001
dTMT B/A	0.13	229	<0.001
dTMT MD A	0.16	229	<0.001
dTMT MD B	0.21	229	<0.001
dTMT ISD A	0.15	229	<0.001

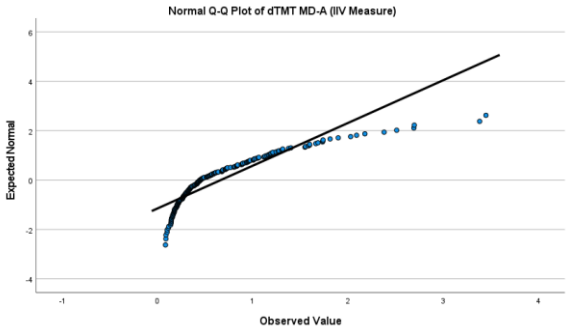
	Statistic	*df*	*p*
dTMT ISD B	0.20	229	<0.001
dTMT rISD A	0.18	229	<0.001
dTMT rISD B	0.21	229	<0.001

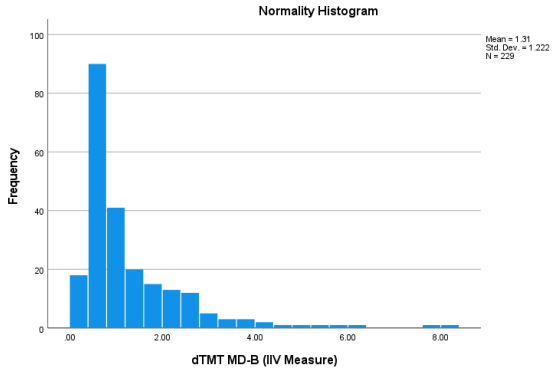
Note. The normality assumption for the correlation analysis was violated for all variables. Variables were measured in time in seconds. Lilliefors Significance Correction.

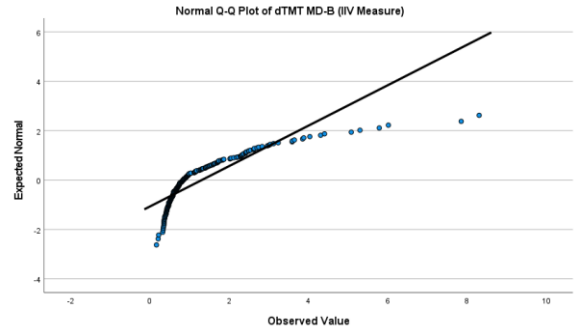
The corresponding histograms and Q-Q plots, Figures A7 through A24, also depict the failed normality of the variables of the second research question below.

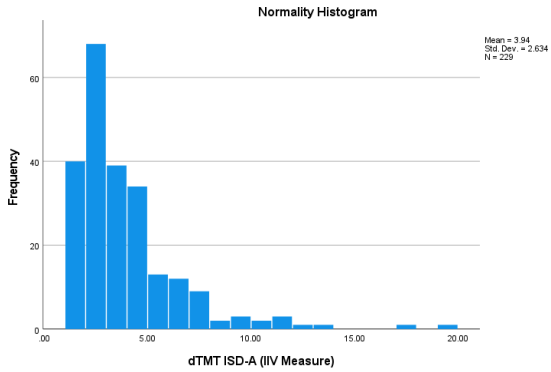


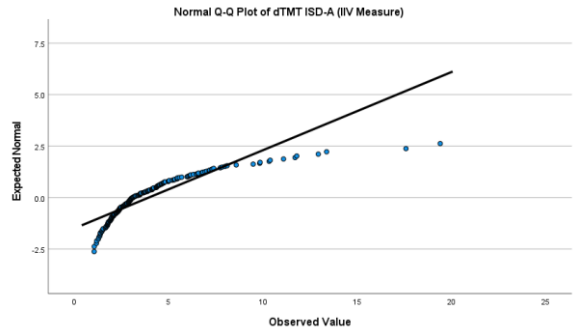

Normality Histogram
Mean = .66
Std. Dev. = .578
N = 229
Frequency
dTMT MD-A (IIV Measure)

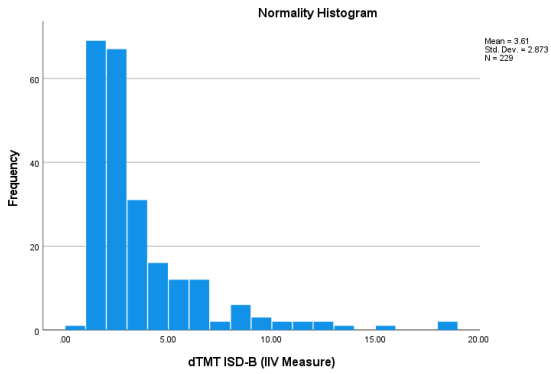

Normal Q-Q Plot of dTMT MD-A (IIV Measure)
Expected Normal
Observed Value

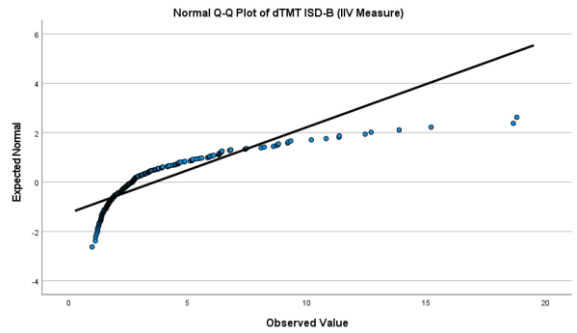

Normality Histogram
Mean = 1.31
Std. Dev. = 1.222
N = 229
Frequency
dTMT MD-B (IIV Measure)

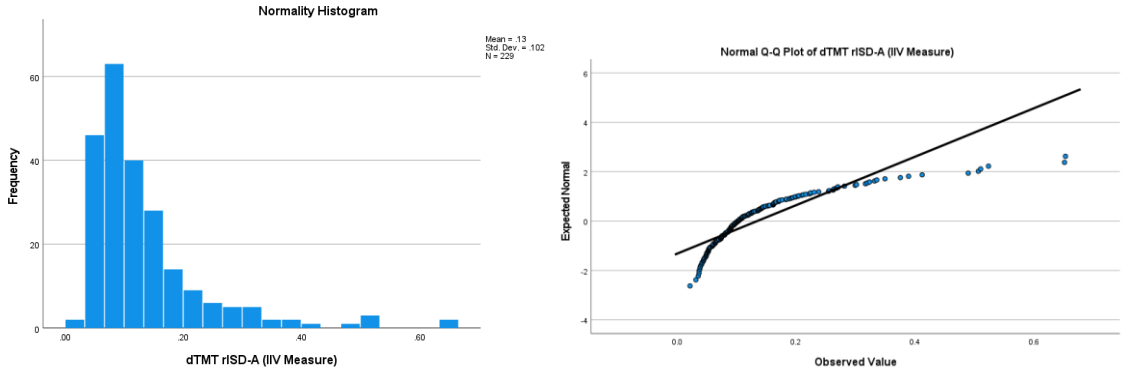

Normal Q-Q Plot of dTMT MD-B (IIV Measure)
Expected Normal
Observed Value

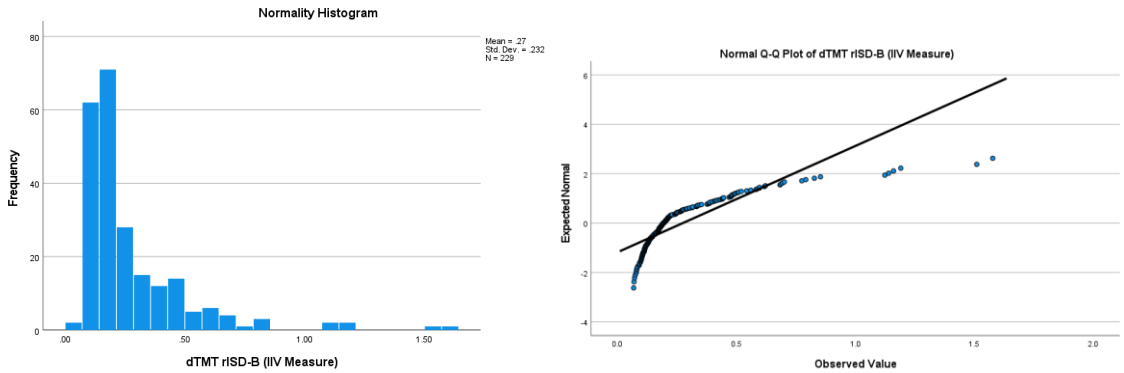

Normality Histogram
Mean = 3.94
Std. Dev. = 2.634
N = 229
Frequency
dTMT ISD-A (IIV Measure)


Normal Q-Q Plot of dTMT ISD-A (IIV Measure)
Expected Normal
Observed Value

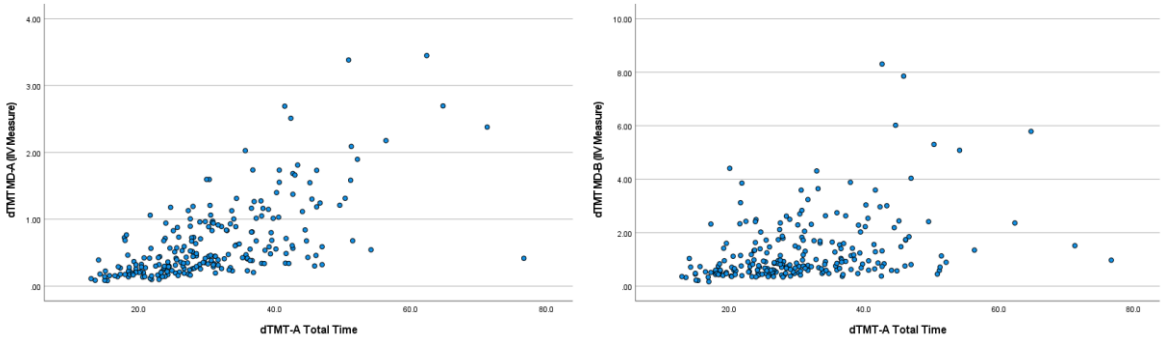

Normality Histogram
Mean = 3.61
Std. Dev. = 2.873
N = 229
Frequency
dTMT ISD-B (IIV Measure)

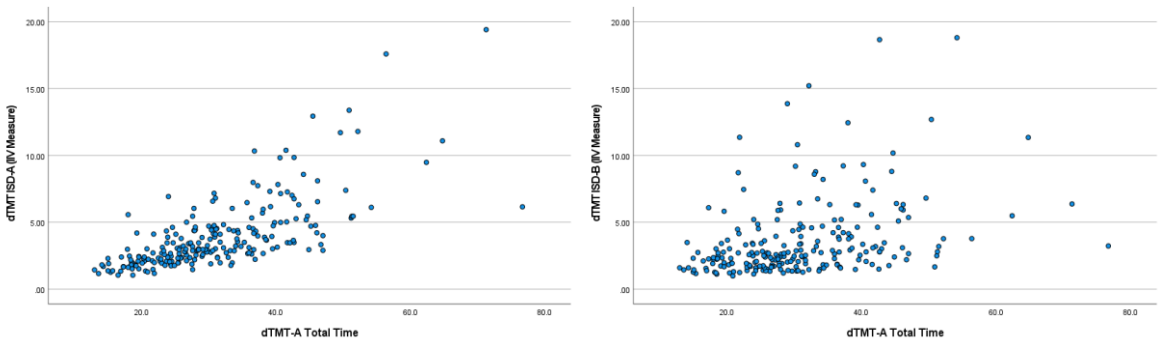

Normal Q-Q Plot of dTMT ISD-B (IIV Measure)
Expected Normal
Observed Value

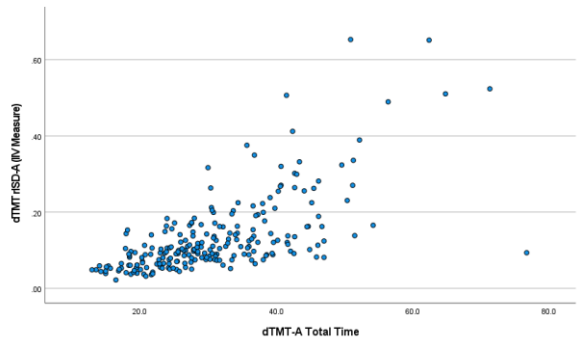


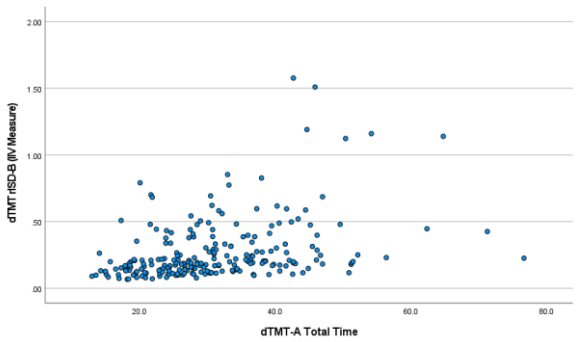


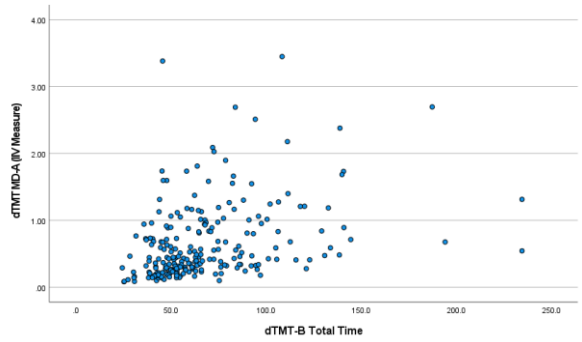
The following scatterplots (Figures A25 through A42) show that the correlational variables failed the linearity assumption as well.

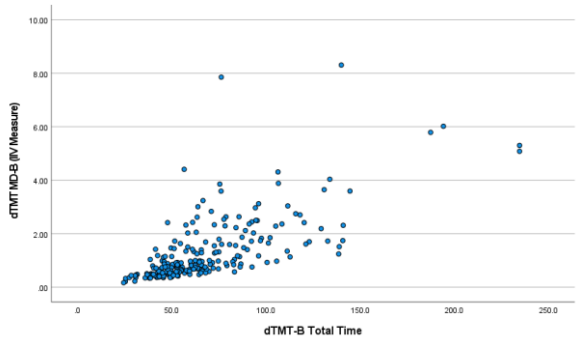


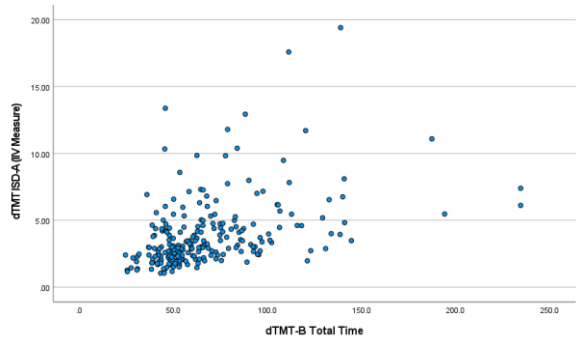


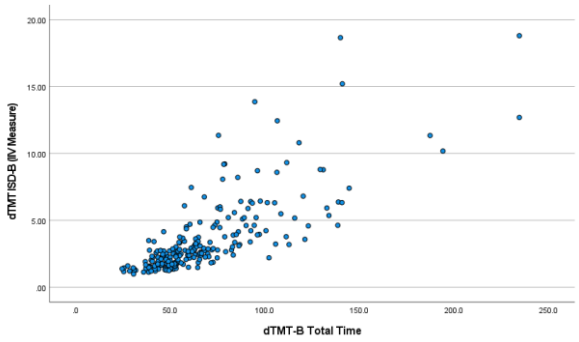

dTMT rISD-A (ft/ Measure)
dTMT-A Total Time

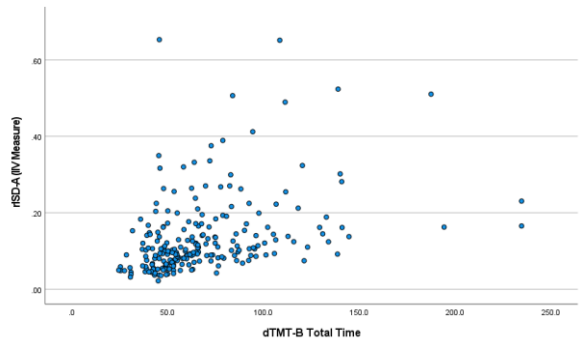

dTMT rISD-B (ft/ Measure)
dTMT-A Total Time

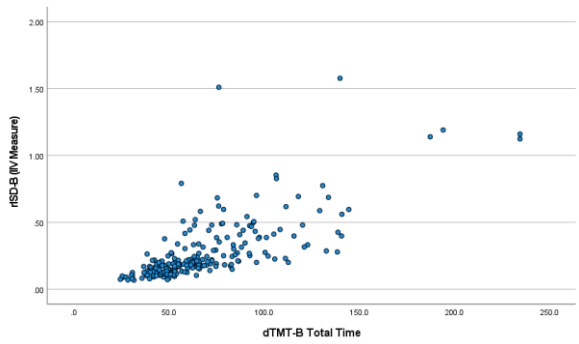

dTMT MD-A (ft/ Measure)
dTMT-B Total Time

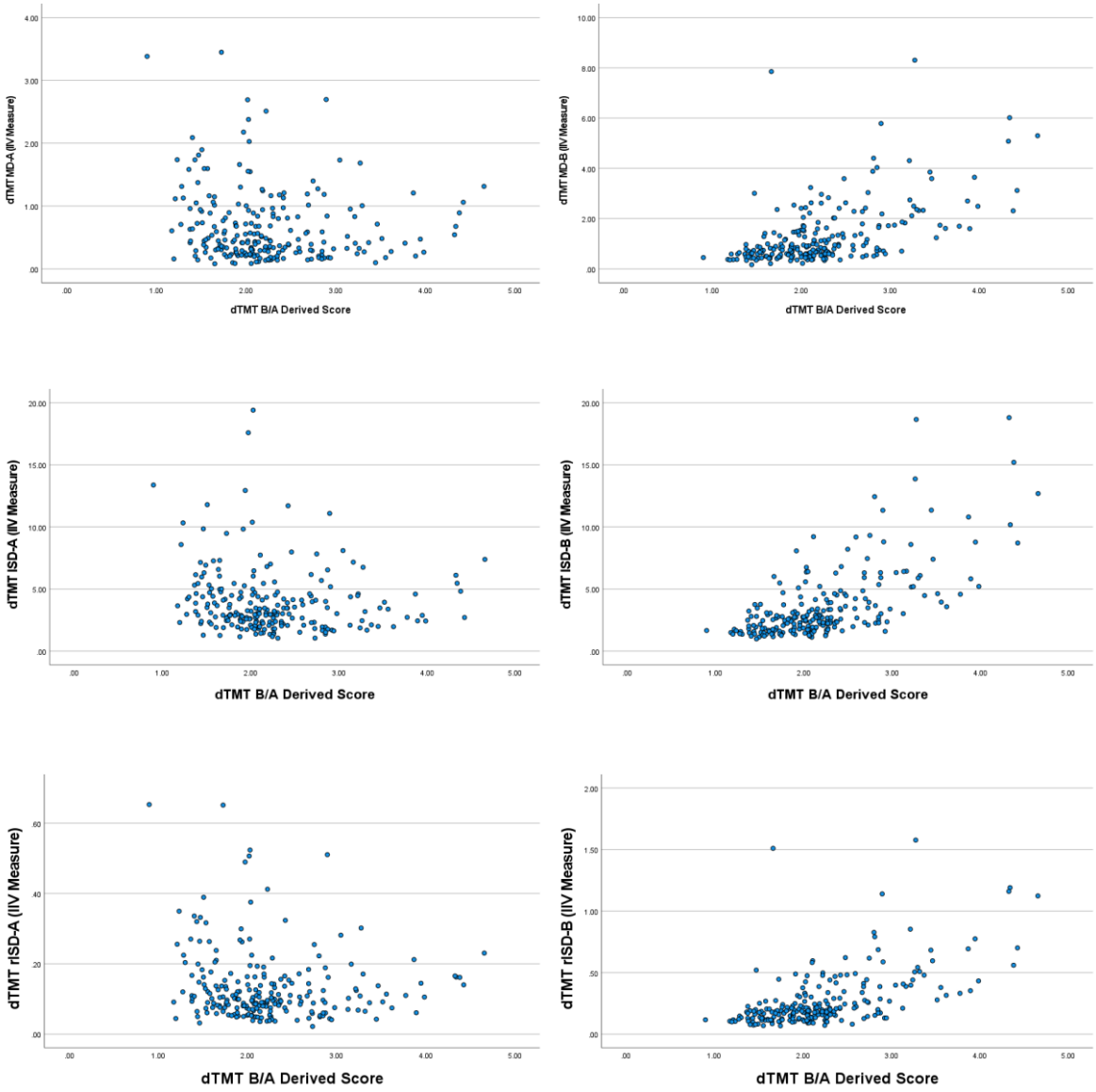

dTMT MD-B (ft/ Measure)
dTMT-B Total Time


dTMT ISD-A (ft/ Measure)
dTMT-B Total Time

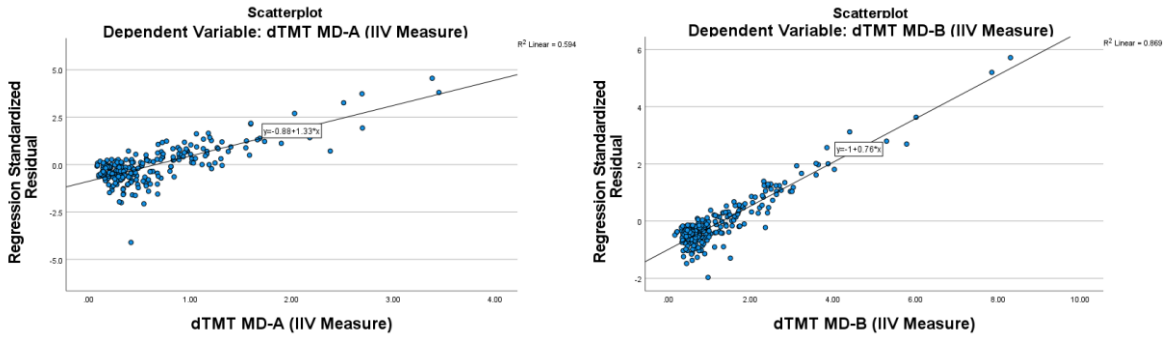

dTMT ISD-B (ft/ Measure)
dTMT-B Total Time

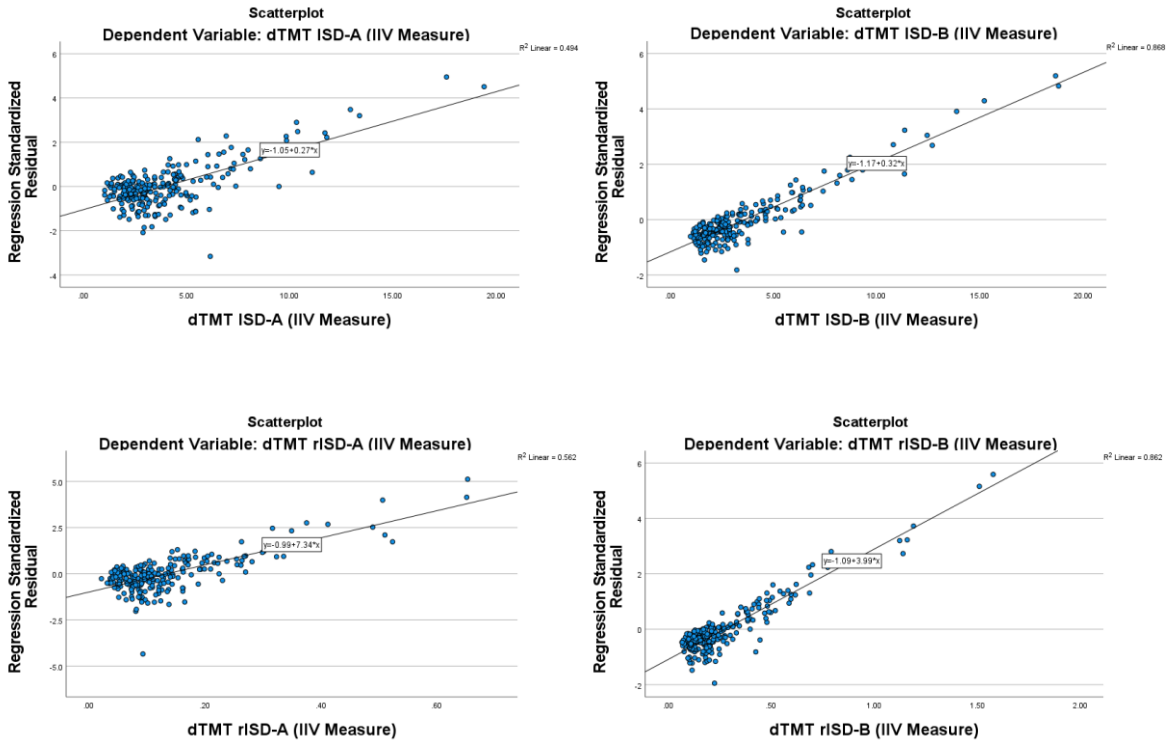

rISD-A (ft/ Measure)
dTMT-B Total Time


rISD-B (ft/ Measure)
dTMT-B Total Time

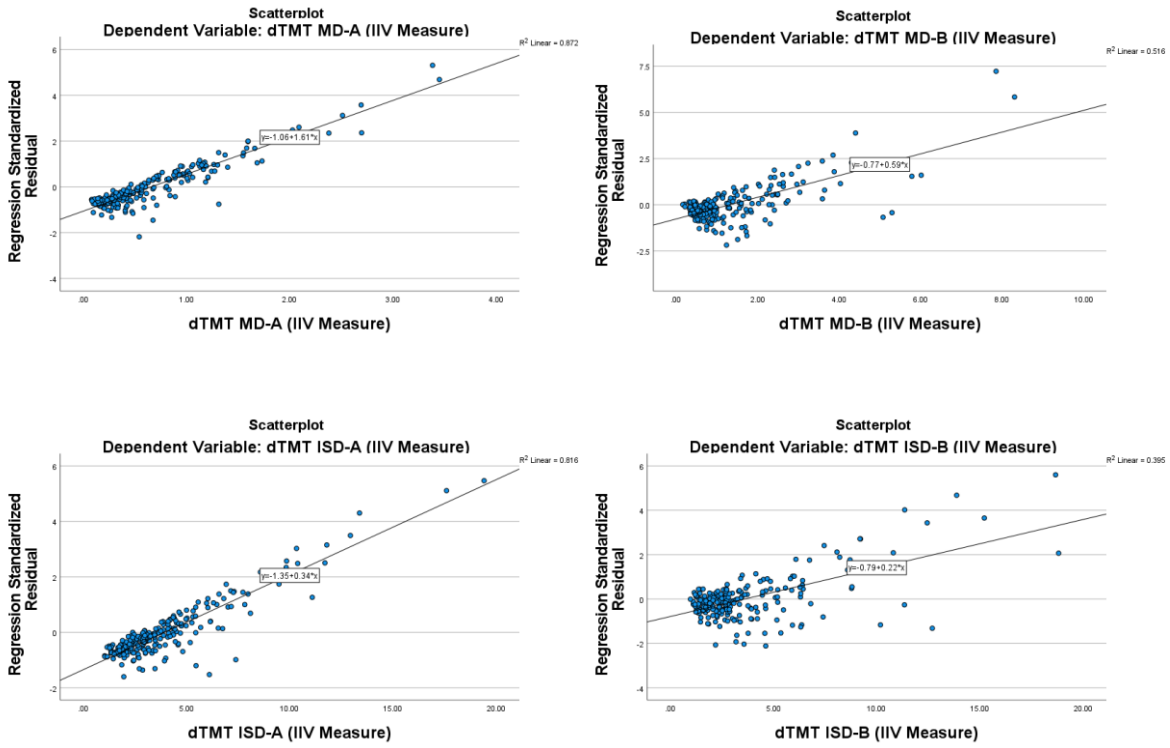


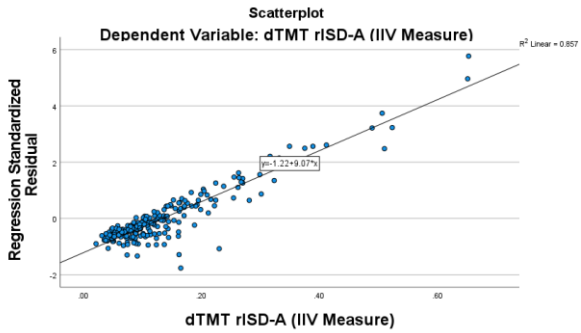
The standardized residual scatterplots, Figures A43 to A48, below show the failed homoscedascity by the IIV measures and dTMT A Total Time involved in the second research question.

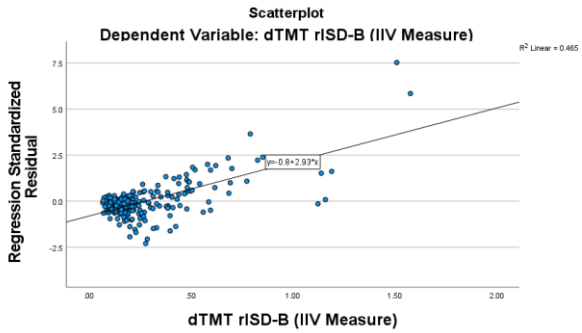




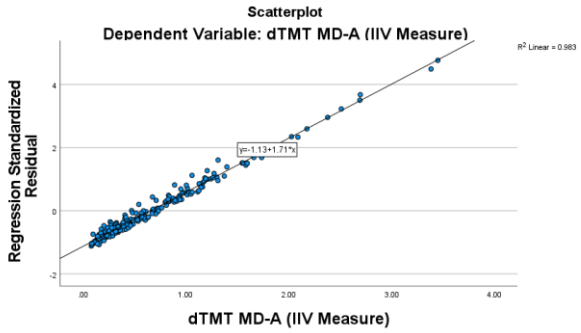
Figures A49 to A54 are scatterplots that show the IIV measures with the dTMT B Total Time failing the homoscedascity assumption requirement.

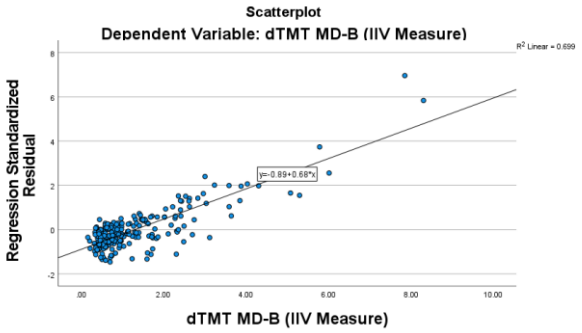


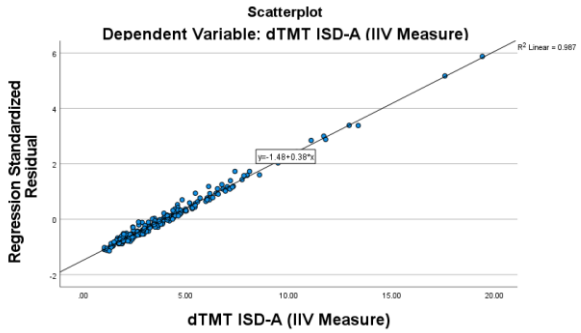


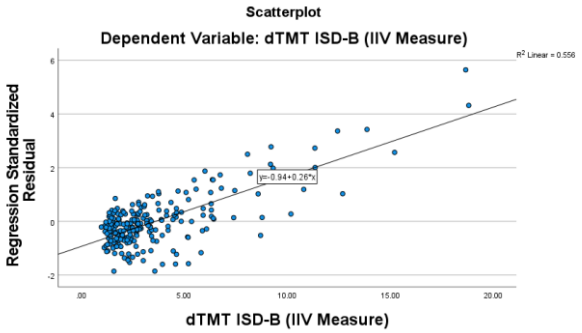


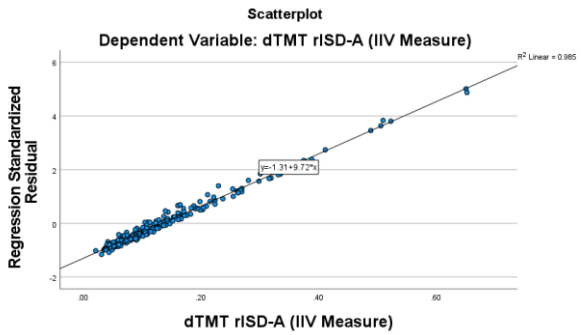
Figures A55 through A60 reflect the IIV measures along with the dTMT B/A Ratio Score. Likewise, they also failed to reflect homoscedascity.

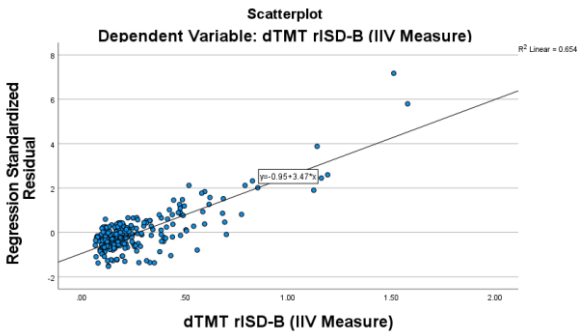












RQ 3: Comparison of IIV Measures Between the HC and ABI Groups

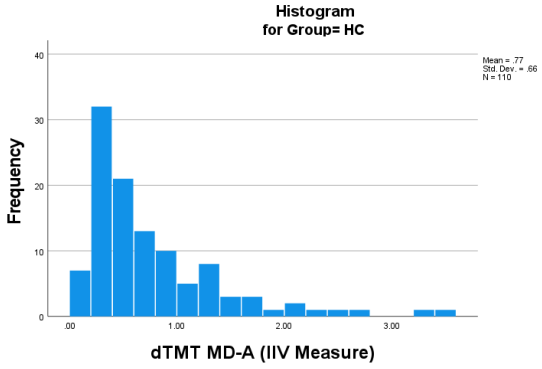
The ‘Tests of Normality’ table A3 below depicts that the normality assumption was violated for the Independent Samples T-Test for the HC and ABI matched groups. This violation is further supported by the histograms and scatterplots (Figures A61 to A76) that follow for the MD and ISD measures.

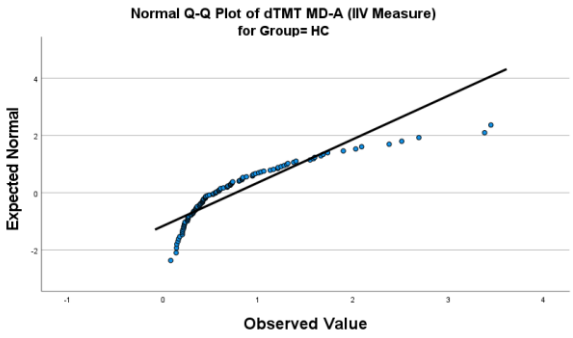
Table A3

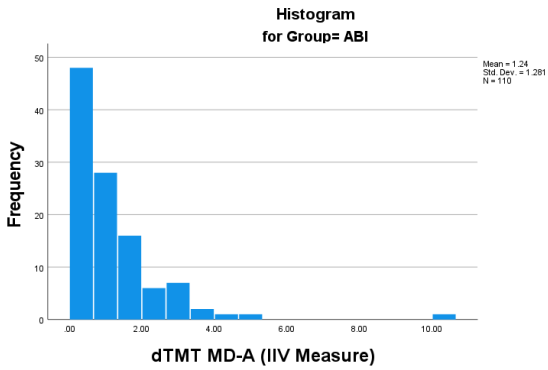
Kolmogorov-Smirnov Test of Normality Between Variables of the dTMT, MD, and ISD in the HC and ABI Populations

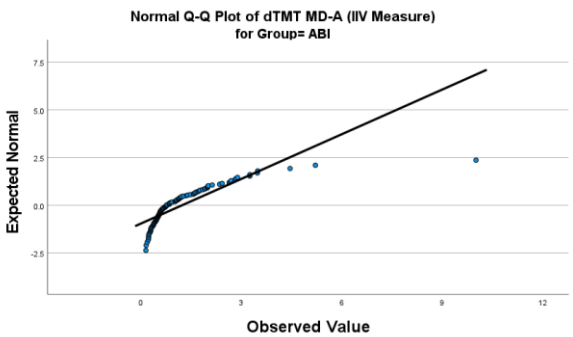
	Statistic	*df*	*p*
HC			
MD A	0.17	110	<0.001
MD B	0.18	110	<0.001
ISD A	0.17	110	<0.001
ISD B	0.22	110	<0.001
ABI			
MD A	0.20	110	<0.001
MD B	0.27	110	<0.001
ISD A	0.17	110	<0.001
ISD B	0.24	110	<0.001

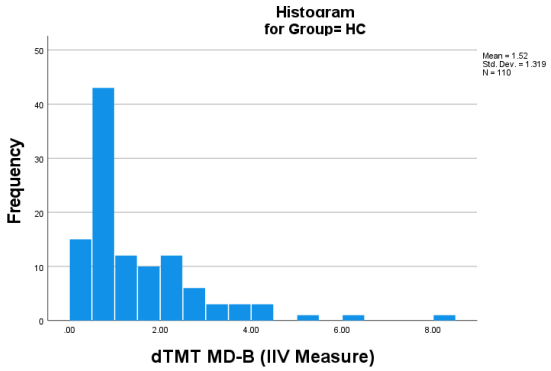
Note. Table A3. The normality assumption for the Independent Samples T-Test was violated for all variables. Lilliefors Significance Correction

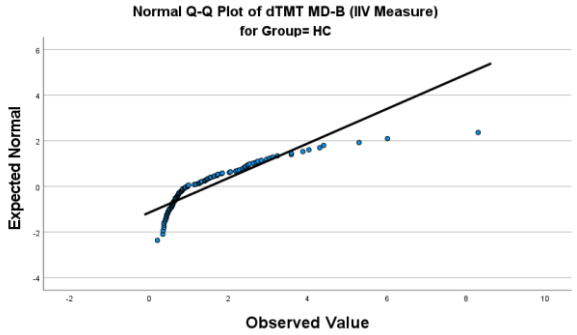


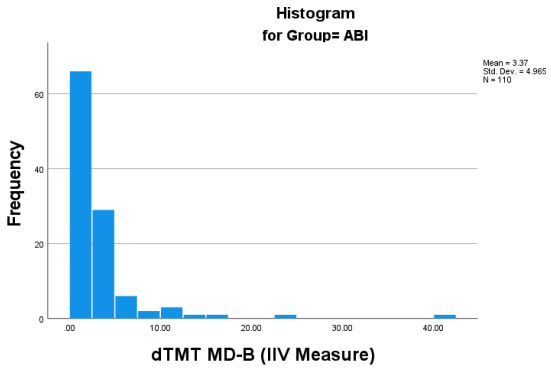


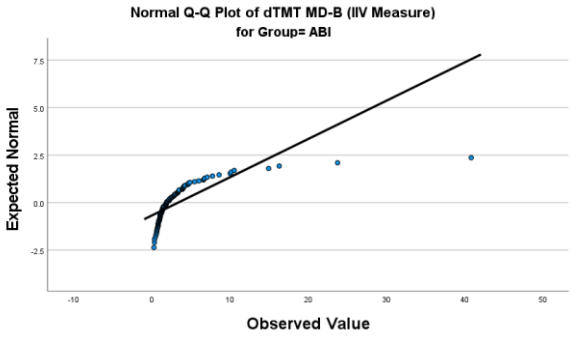


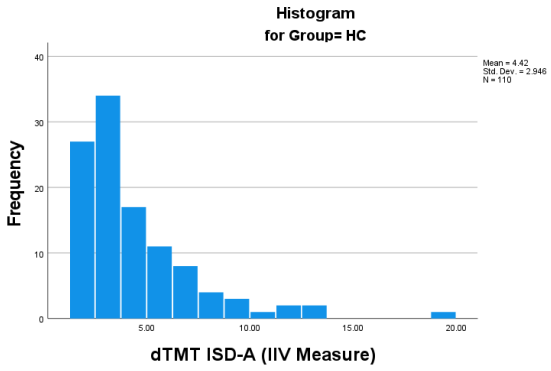



Histogram
for Group= HC
Mean = 1.52
Std. Dev. = 1.319
N = 110
Frequency
dTMT MD-B (IIV Measure)

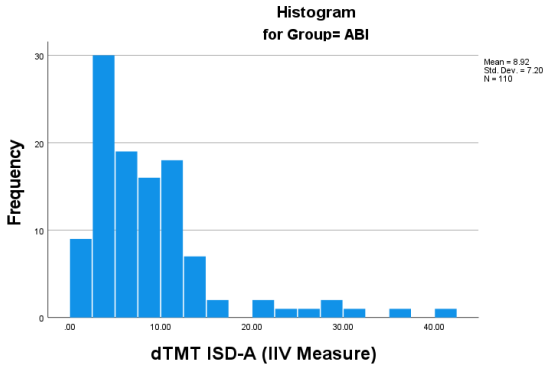

Normal Q-Q Plot of dTMT MD-B (IIV Measure)
for Group= HC
Expected Normal
Observed Value

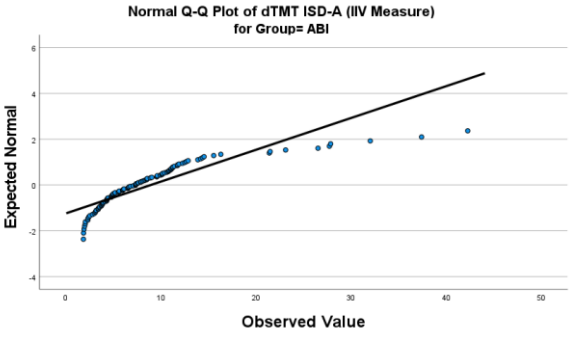

Histogram
for Group= ABI
Mean = 3.37
Std. Dev. = 4.965
N = 110
Frequency
dTMT MD-B (IIV Measure)

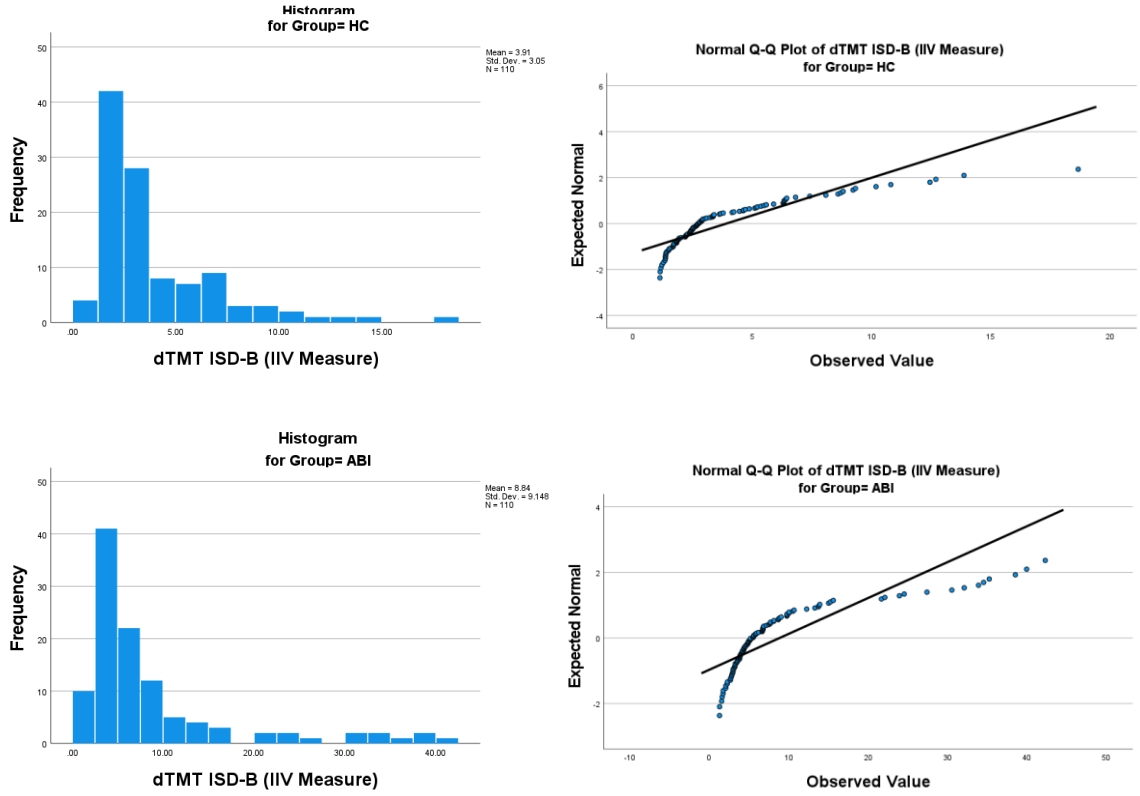

Normal Q-Q Plot of dTMT MD-B (IIV Measure)
for Group= ABI
Expected Normal
Observed Value


Histogram
for Group= HC
Mean = 4.42
Std. Dev. = 2.946
N = 110
Frequency
dTMT ISD-A (IIV Measure)


Normal Q-Q Plot of dTMT ISD-A (IIV Measure)
for Group= HC
Expected Normal
Observed Value


Histogram
for Group= ABI
Mean = 8.92
Std. Dev. = 7.20
N = 110
Frequency
dTMT ISD-A (IIV Measure)


Normal Q-Q Plot of dTMT ISD-A (IIV Measure)
for Group= ABI
Expected Normal
Observed Value



The violation of the homogeneity assumption is seen below in the Test of Homogeniety of Variance, Table A4.

Table A4

Test of Homogeneity of Variance for the MD and ISD IIV Measures

		Levene Statistic	*df1*	*df2*	*p*
MD A	Based on Mean	12.25	1	218	<0.001
	Based on Median	6.88	1	218	0.009
	Based on Median and with adjusted *df*	6.88	1	154.61	0.010
	Based on trimmed mean	9.54	1	218	0.002
MD B	Based on Mean	15.50	1	218	<0.001
	Based on Median	8.75	1	218	0.003
	Based on Median and with adjusted *df*	8.75	1	121.31	0.004
	Based on trimmed mean	10.50	1	218	.001
ISD A	Based on Mean	25.66	1	218	<0.001
	Based on Median	20.29	1	218	<0.001
	Based on Median and with adjusted *df*	20.29	1	149.84	<0.001
	Based on trimmed mean	21.89	1	218	<0.001
ISD B	Based on Mean	36.22	1	218	<0.001
	Based on Median	16.91	1	218	<0.001
	Based on Median and with adjusted *df*	16.91	1	131.01	<0.001
	Based on trimmed mean	25.76	1	218	<0.001

Note. Table A4 shows that the variables failed the homogeneity of variance assumption.

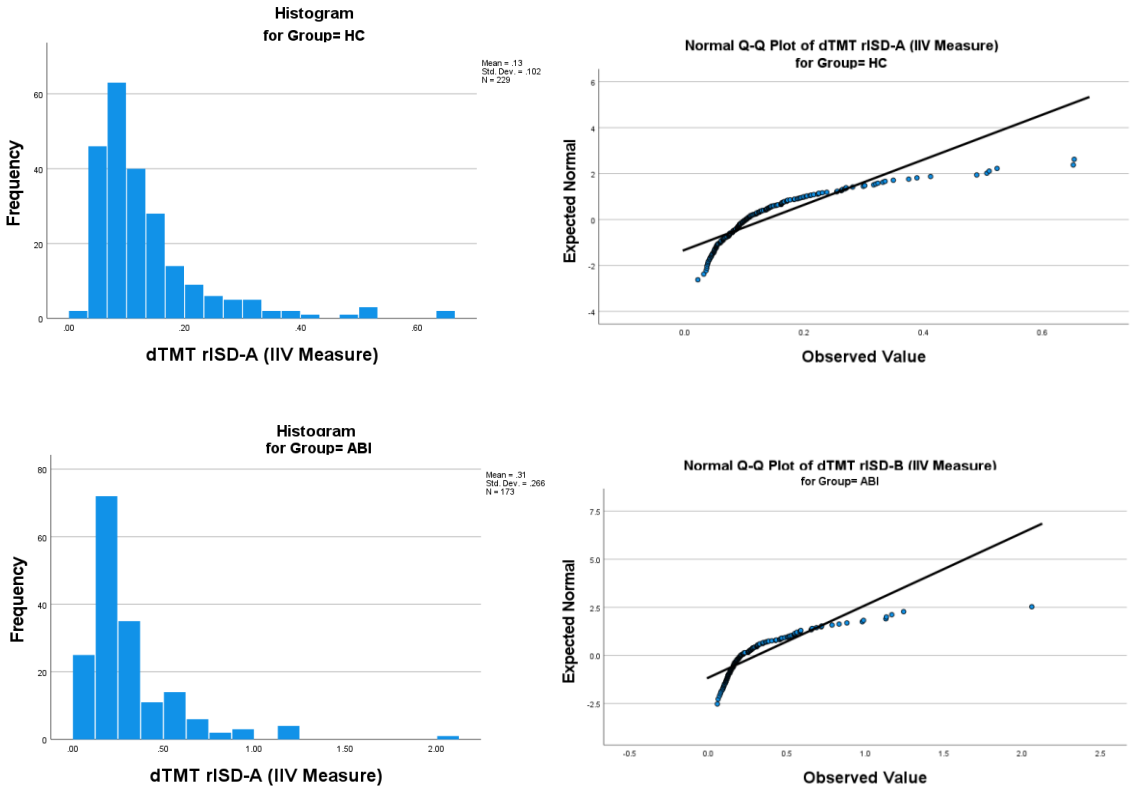
A second Independent Samples T-Test was conducted with the whole sample and the rISD measures. The normality assumption was violated as seen by the supporting Table A5, histograms, and scatterplots (Figures A76 to A83).

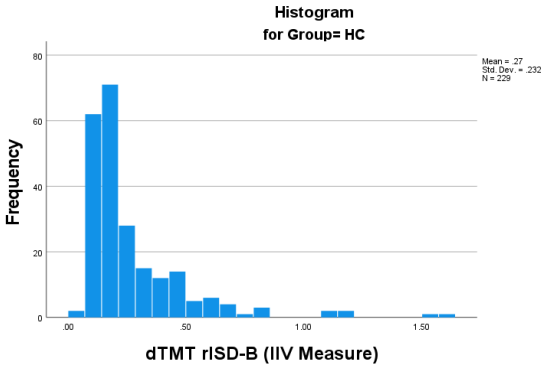
Table A5

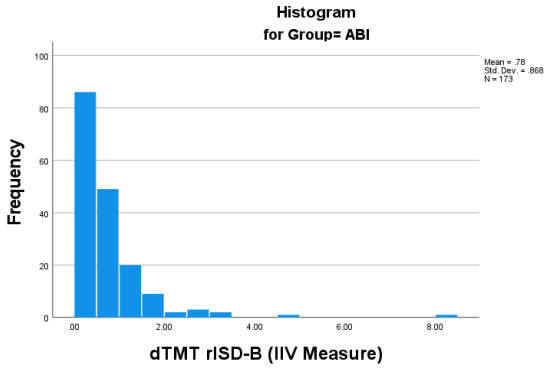
Kolmogorov-Smirnov Test of Normality Between Variables of the dTMT and rISD in the HC and ABI Groups

	Statistic	*df*	*p*
HC			
rISD A	0.18	229	<0.001
rISD B	0.21	173	<0.001
ABI			
rISD A	0.20	229	<0.001
rISD B	0.23	173	<0.001

Note. Table A5. The normality assumption for the Independent Samples T-Test was violated for all variables. Lilliefors Significance Correction.







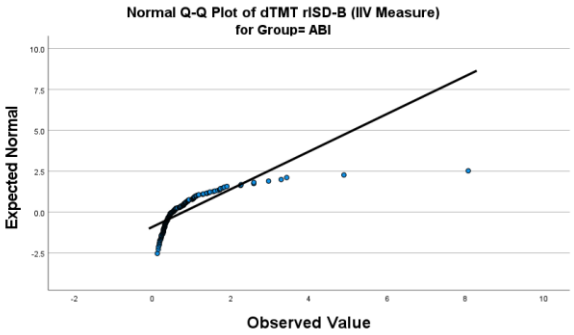


Table A6 shows that the Test of Homogeneity of Variance for the rISD measures also failed the assumption requirements.

Table A6

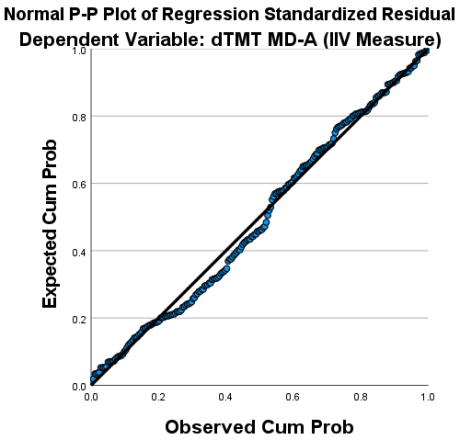
Test of Homogeneity of Variance for the rISD IIV Variables

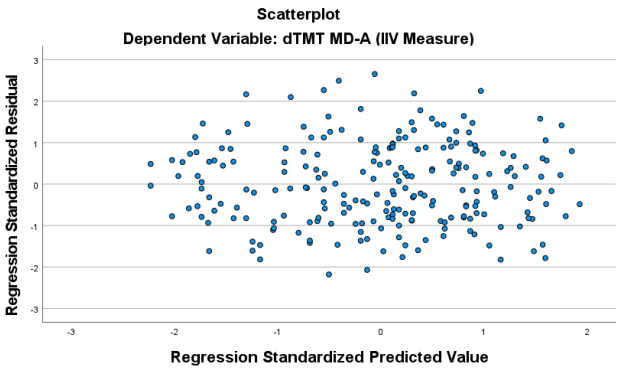
		Levene Statistic	*df1*	*df2*	*p*
rISD A	Based on Mean	60.98	1	400	<0.001
	Based on Median	33.95	1	400	<0.001
	Based on Median and with adjusted df	33.95	1	233.92	<0.001
	Based on trimmed mean	47.11	1	400	<0.001
rISD B	Based on Mean	49.10	1	400	<0.001
	Based on Median	31.68	1	400	<0.001
	Based on Median and with adjusted df	31.68	1	202.26	<0.001
	Based on trimmed mean	39.03	1	400	<0.001

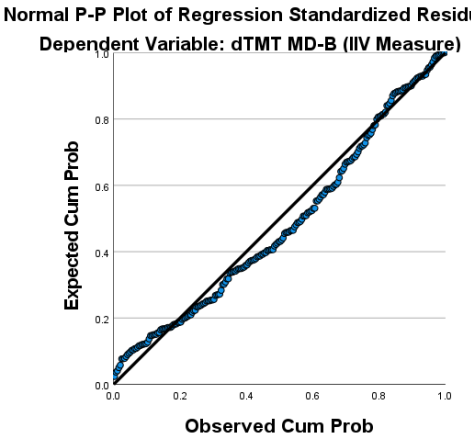
Note. Table A6 shows that the rISD A and rISD B variables fail the assumption for homogeneity of variance.

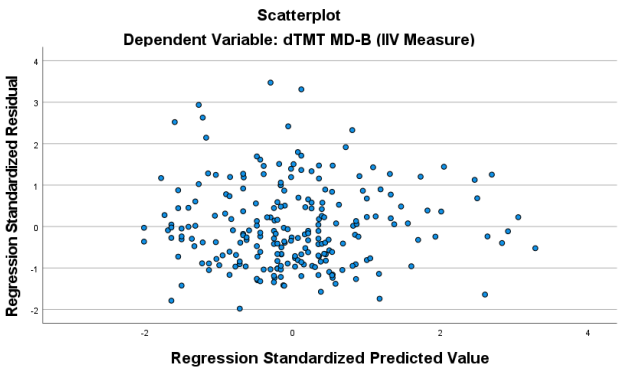
RQ 4: Multiple Regression Analyses to Analyze Which Predictor Variables Influence IIV in HCs and ABIs

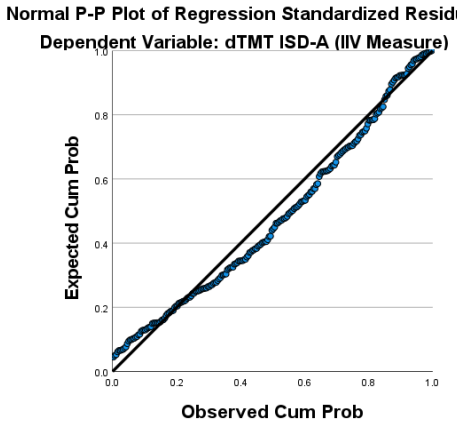
There were two multiple regression models analyzed in the fourth research question. The first used four predictor variables (Handedness, Age, Gender, and Education Level) and below are the residuals plotted of the model. All six assumptions were met in this model (Figures A86 to A97).

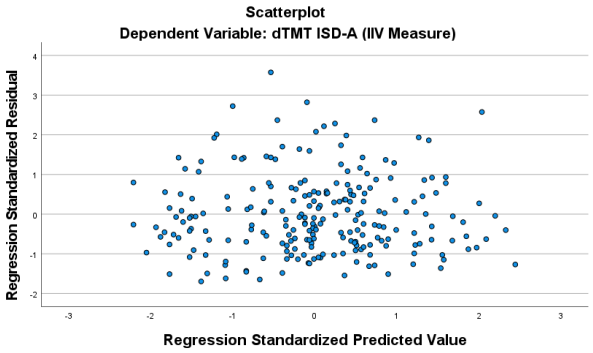


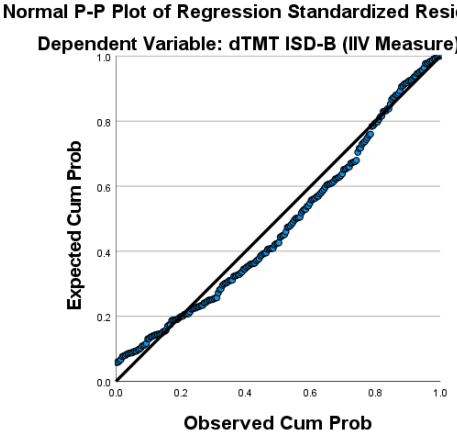


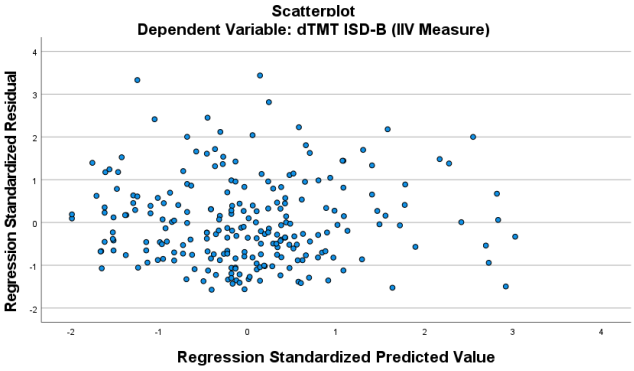


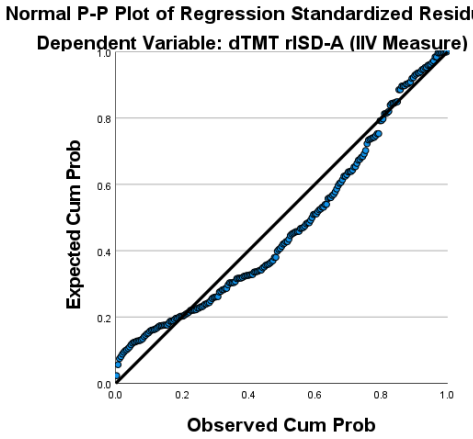


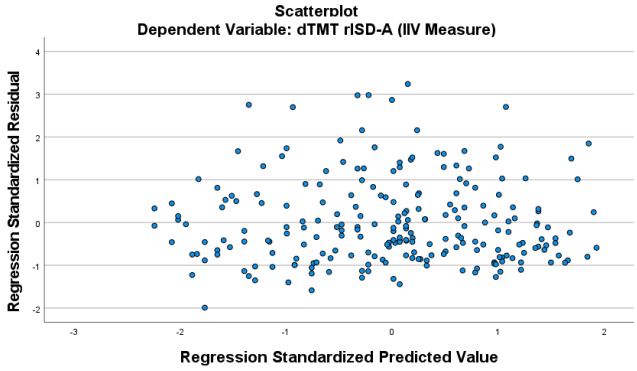

Normal P-P Plot of Regression Standardized Residual
Dependent Variable: dTMT ISD-A (IIV Measure)
Expected Cum Prob
Observed Cum Prob

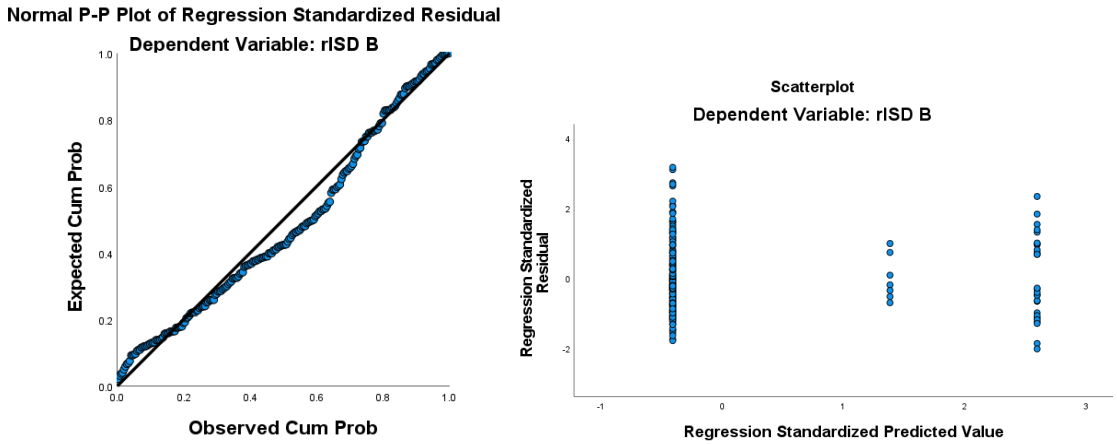

Scatterplot
Dependent Variable: dTMT ISD-A (IIV Measure)
Regression Standardized Residual
Regression Standardized Predicted Value


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: dTMT ISD-B (IIV Measure)
Expected Cum Prob
Observed Cum Prob

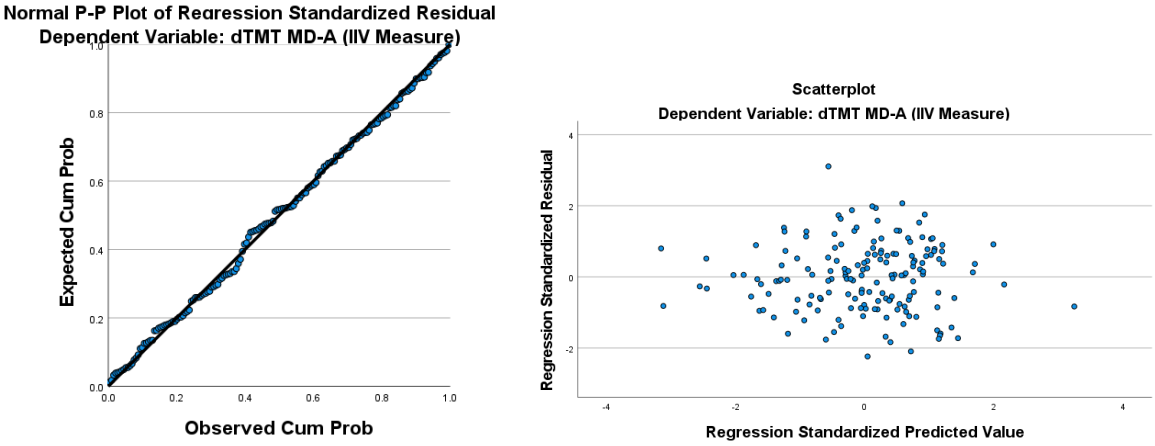

Scatterplot
Dependent Variable: dTMT ISD-B (IIV Measure)
Regression Standardized Residual
Regression Standardized Predicted Value

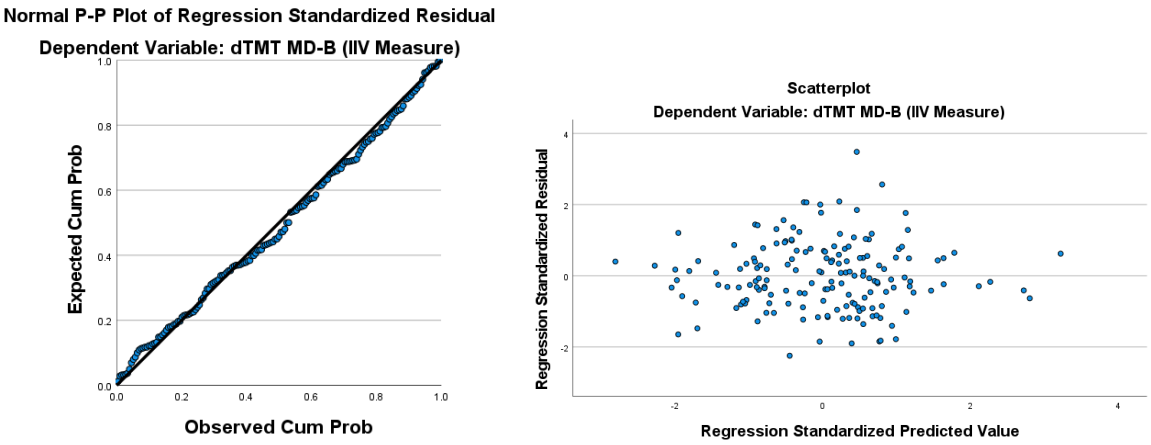

Normal P-P Plot of Regression Standardized Residual
Dependent Variable: dTMT rISD-A (IIV Measure)
Expected Cum Prob
Observed Cum Prob

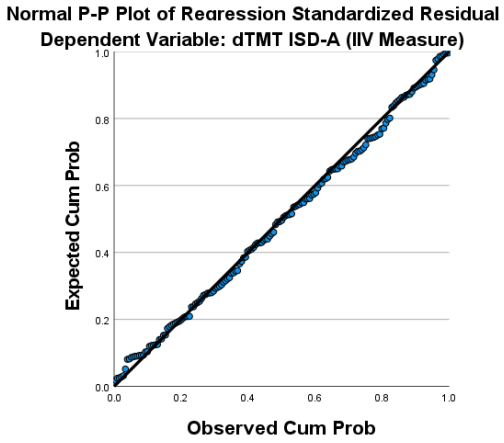

Scatterplot
Dependent Variable: dTMT rISD-A (IIV Measure)
Regression Standardized Residual
Regression Standardized Predicted Value

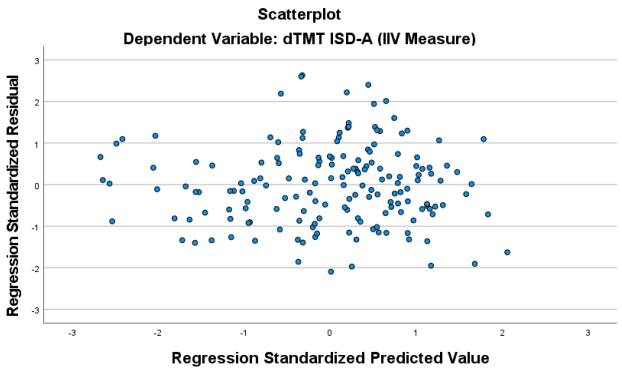


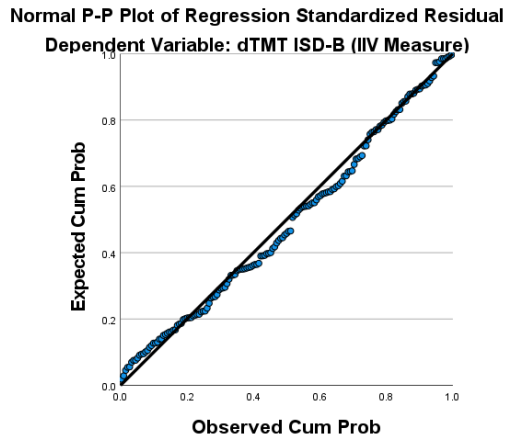
The second multiple regressions used eight predictor variables (Handedness, Age, Gender, Education Level, ABI etiology, side of ABI, affected binocular visual field, and Time Since Recent ABI) and are the residuals plotted of the model. All six assumptions were met in this model as seen in Figures A98 to A109.

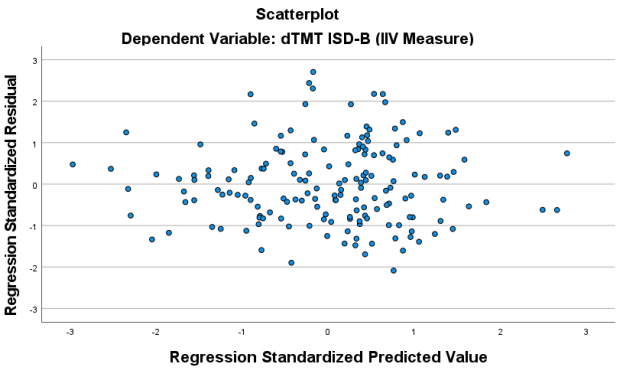


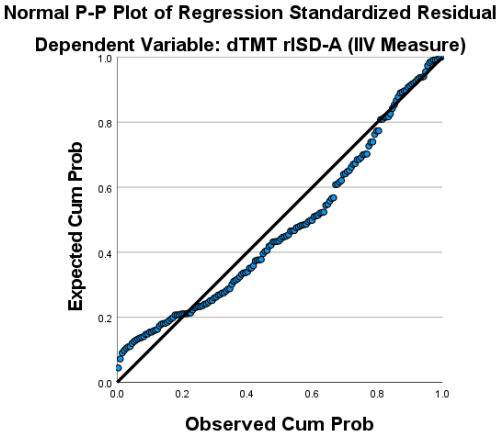


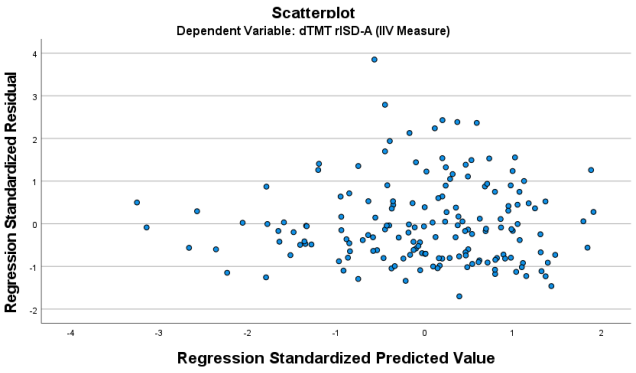

Normal P-P Plot of Regression Standardized Residual
Dependent Variable: dTMT ISD-A (IIV Measure)
Expected Cum Prob
Observed Cum Prob

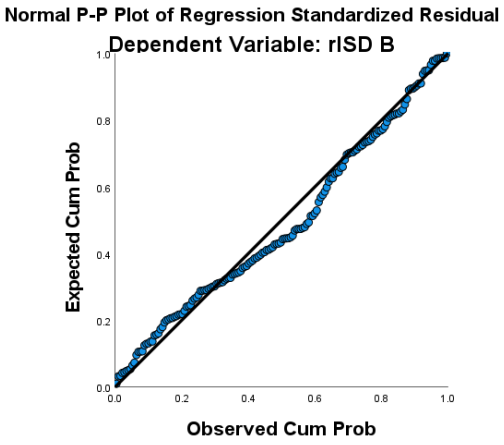

Scatterplot
Dependent Variable: dTMT ISD-A (IIV Measure)
Regression Standardized Residual
Regression Standardized Predicted Value

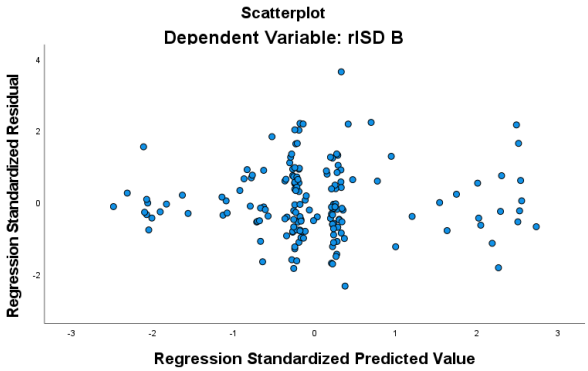

Normal P-P Plot of Regression Standardized Residual
Dependent Variable: dTMT ISD-B (IIV Measure)
Expected Cum Prob
Observed Cum Prob


Scatterplot
Dependent Variable: dTMT ISD-B (IIV Measure)
Regression Standardized Residual
Regression Standardized Predicted Value


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: dTMT rISD-A (IIV Measure)
Expected Cum Prob
Observed Cum Prob


Scatterplot
Dependent Variable: dTMT rISD-A (IIV Measure)
Regression Standardized Residual
Regression Standardized Predicted Value


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: rISD B
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Scatterplot
Dependent Variable: rISD B
Regression Standardized Residual
Regression Standardized Predicted Value
4
2
0
-2
-3
-2
-1
0
1
2
3