PROTOCOL DCT032: STATISTICAL ANALYSIS PLAN

Digital Cognition Technologies, Inc.

210 Bear Hill Road, Suite 300

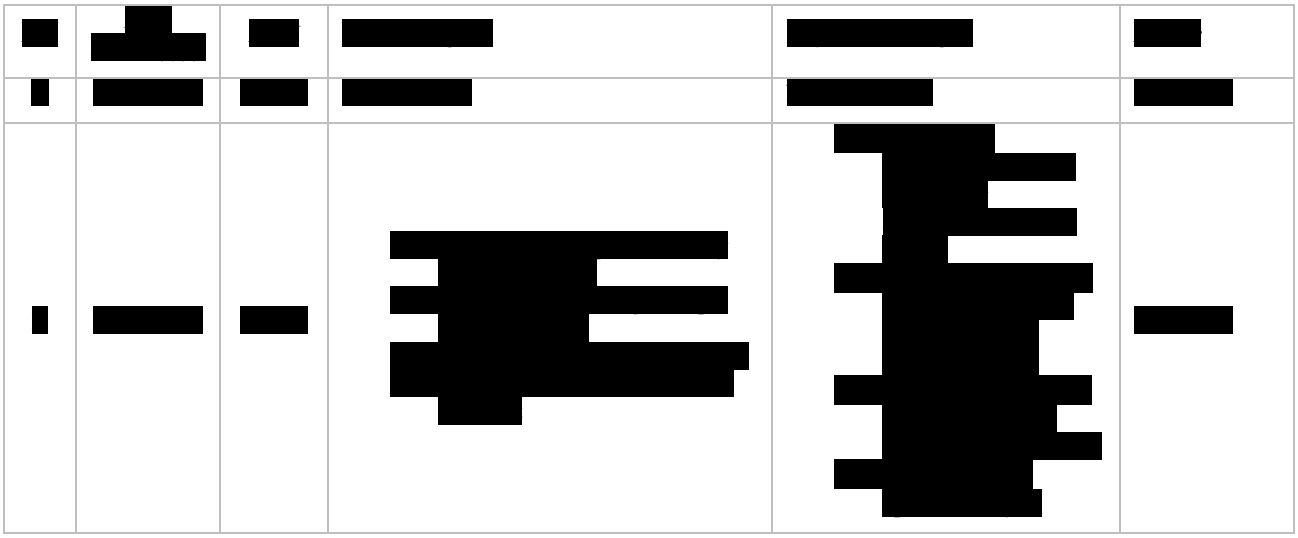
Waltham, MA 02451

Protocol Identifying No. DCT032: September 21, 2017

Version: B

Authors: [redacted]

[redacted]



List of Abbreviations

Abbreviation	Definition
AE	Adverse Event
CI	Confidence Interval
H_0	Null Hypothesis
H_1	Alternative Hypothesis
ITT	Intention-to-Treat
NPA	Negative Percent Agreement
MMSE	Mini-Mental State Examination
PPA	Positive Percent Agreement
QC	Quality Control
SAP	Statistical Analysis Plan

Contents

1 Introduction 4
1.1 Purpose of Statistical Analysis Plan 4
2 Study Objectives 4
2.1 Overview 4
2.2 Primary Objective 4
2.3 Secondary Objectives 4
3 Study Design 5
3.1 Overview 5
3.2 Test Device 5
3.3 Sample Size 5
3.4 Populations 6
3.4.1 Intention-to-Treat Population (ITT) 6
3.4.2 Evaluable Population 6
3.4.3 Safety Population 6
4 Effectiveness and Safety Endpoints 6
4.1 Primary Effectiveness Endpoint 6
4.2 Secondary Effectiveness Endpoints 6
4.3 Other Effectiveness Endpoints 6
4.4 Safety Endpoints 7
5 Statistical Methods 7
5.1 Statistical Analyses 7
5.1.1 Primary Effectiveness Analysis (DCTclock/MMSE Agreement) 7
5.1.2 Secondary Effectiveness Analysis (Test/Re-Test Reliability) 8
5.1.3 Additional Effectiveness Analyses 10
5.1.4 Safety Analysis 11
5.1.5 Additional Summaries 11
5.2 Data Handling Procedures 11
5.2.1 Data Collection, Entry, and Quality Control 11

5.2.2 Missing Data 12
5.2.3 Rounding 12
5.3 Statistical Software 12
6 References 13

1 Introduction

1.1 Purpose of Statistical Analysis Plan

The purpose of this SAP is to outline the necessary statistical methods and procedures for protocol identifying number DCT032. The plan outlined will cover the statistical methodology for the final report only, as there are no planned interim analyses.

2 Study Objectives

2.1 Overview

To demonstrate the safety and effectiveness of DCTclock as an adjunctive tool for use by clinicians to evaluate cognitive function in adults aged 55-95.

2.2 Primary Objective

Examine the agreement between DCTclock classifications and a reference standard, the Mini-Mental State Exam (MMSE).

2.3 Secondary Objectives

- Determine the test/re-test reliability of DCTclock
- Determine the construct validity of DCTclock via comparison with traditional paper and pencil neuropsychological tests
- Evaluate the safety of DCTclock by characterizing the incidence of serious device-related adverse events (AE)

3 Study Design

3.1 Overview

This study uses a paired design in which each subject undergoes blinded testing by separate testers administering DCTclock and a battery of reference standard tests to determine cognitive state. Each subject undergoes a first set of testing with repeated testing after a one to four week interval. Every subject will undergo testing with DCTclock, MMSE, Montreal Cognitive Assessment (MoCA), Digit Span and Reliable Digit Calculation, and Trail Making Test parts A and B. Agreement between the DCTclock and the MMSE will be conducted by using the first test. Test-retest reliability of both DCTclock and MMSE will be conducted by comparing performance on the first and second test. 500 subjects will be recruited to participate with the anticipation that 400 will be available for analysis. A number of participants will also be administered the Benton Judgement of Line Orientation, Rey Complex Figure, Delis-Kaplan Executive Function System Verbal Fluency, Block Designs, Symbol Digit Modalities, and the Geriatric Depression Scale.

3.2 Test Device

DCTclock, a digitized version of the standard pen and paper neuropsychological clock drawing test, is a non-invasive, computer-based cognitive assay. The test involves participants drawing two clock faces on a piece of paper with a digital pen that precisely tracks and records drawing behavior. The positional data generated during this assessment is then analyzed by proprietary algorithms that evaluate hundreds of features captured in the pen stroke information. By comparing test results to normative data, the system then determines an overall 0-100 score as well as a categorical output: inside normal limit, indeterminate, outside normal limits. The test also provides a detailed breakdown of performance on the various cognitive processes evaluated during the test. The intended users of the device are individuals aged 55-95 years old.

3.3 Sample Size

The sample size for this study includes 429 participants with a non-missing DCTclock and MMSE on the first visit. The Montreal Cognitive Assessment (MoCA), Reitan Trails A (TRA), Reitan Trails B, and the Wechsler Adult Intelligence Scale Digit Span (WAIS-IV DS) were administered to all patients in the study, with small amounts of missing tests due to documented protocol deviations. Other NP tests, including the Benton Judgement of Line Orientation (BJLO), the Delis-Kaplan Executive Function System Verbal Fluency Test (DKEFS VF), the WAIS-IV Block Design (WAIS-IV BD), the Rey Complex Figure Test (Rey), and Symbol Digit Modalities (SDM) were administered to a subset of roughly 190 patients, again with a small number of missing tests.

3.4 Populations

3.4.1 Intention-to-Treat Population (ITT)

The Intention-to-treat population is defined as all enrolled patients.

3.4.2 Evaluable Population

The evaluable population is defined as all enrolled participants undergoing cognitive testing with both DCTclock and MMSE and who are not performing sub-optimally (as determined by the Reliable Digit Score). This is the primary analysis set for effectiveness.

3.4.3 Safety Population

The safety population is defined as all enrolled participants undergoing cognitive testing with DCTclock. This is the primary analysis set for safety.

4 Effectiveness and Safety Endpoints

4.1 Primary Effectiveness Endpoints

DCTclock score, MMSE score and MoCA tests scores on the primary endpoint population. The goal is to evaluate the non-inferiority of DCTclock and MMSE assessments of impaired, indeterminate, and unimpaired status at participants' first visit. This will be achieved by comparing the agreement between DCTclock and MoCA classifications with the agreement between MMSE and MoCA classifications.

4.2 Secondary Effectiveness Endpoints

DCTclock scores, MMSE scores and assessment of DCTclock and MMSE impaired and unimpaired status at each of two time points. The secondary analysis will assess the comparability of DCTclock test/re-test reliability to that of MMSE. Quadratic weighted Cohen's kappa statistics for both DCTclock and MMSE scores will be calculated, along with PPA, NPA, percent agreement for impaired, indeterminate, and unimpaired. Unweighted Deming regression, linear regression, and linear rank regression intercept and slope estimates with 95% confidence intervals will be calculated, as well as a chi-square statistic for each pair of DCTclock scores and MMSE scores.

4.3 Other Effectiveness Endpoints

DCTclock scores, MMSE scores and neuropsychology battery scores. The goal is to assess construct validity of DCTclock by comparing scores from the repeated administration of DCTclock to the repeated administration of a battery of neuropsychological tests, to characterize the psychometric properties of DCTclock and to compare those properties to those of MMSE. Correlation coefficients will be calculated between DCTclock scores and scores from a battery of neuropsychological tests. Sensitivity analyses will be conducted on the primary and secondary effectiveness endpoints, varying the threshold for MMSE classification.

4.4 Safety Endpoints

Incidence of serious device-related AEs.

5 Statistical Methods

5.1 Statistical Analyses

5.1.1 Primary Effectiveness Analysis (DCTclock/MMSE Agreement)

1. Primary Analysis of the Primary Endpoint

The primary effectiveness analysis will be carried out using quadratic weighted kappa scores. A two-one-sided test (TOST) approach will be conducted, comparing the agreement between the MMSE and DCTclock scores with MoCA scores. The statistical considerations for conducting a TOST analysis include determining an acceptable level of difference in kappa scores (called δ from this point forward). Typically, the following interpretations hold for Cohen's weighted kappa:

0 = Chance agreement
0.10-0.20 = Slight agreement
0.21-0.40 = Fair agreement
0.41-0.60 = Moderate agreement
0.61-0.80 = High agreement
0.81-0.99 = Near perfect agreement
1 = Perfect Agreement

It is expected that the kappa will be at least 0.30 between the DCTclock and MoCA as well as between the MMSE and MoCA. With this in mind, it's safe to say that it's expected that both kappa statistics will be ≥ 0.21 and would require a maximum difference of 0.20 for the interpretations to differ (based on the scoring system above). Thus, a δ of 0.20 was chosen for this analysis.

To show non-inferiority using a δ of 0.20, the lower limit of a (1-2α)x100% confidence interval for the difference in kappa statistics must be greater than -0.20. The (1-2α)x100% confidence interval is calculated as follows:

$\kappa_{DCTclock} - \kappa_{MMSE} \pm 1.645*SE_{Difference}$ where $SE_{Difference}$ is the standard error of the difference in kappa statistics and will be calculated by finding the bootstrapped variance of the difference in kappa statistics, identical to the method proposed by Vanbelle (2008).

Based on research in the literature (Nasreddine, 2005), MoCA values ≥ 26 are considered to be cognitively healthy, while values of < 24 are almost always considered to be cognitively impaired. Thus, MoCA scores < 24 were chosen to represent the Impaired MoCA..

Table I. 3x3 table used for calculating κ between MMSE/DCTclock and MoCA

		MoCA			
		Impaired	Indeterminate/Intermediate	Unimpaired	**Total**
DCTclock MMSE	Impaired	a	b	c	a+b+c
	Indeterminate/Intermediate	d	e	f	d+e+f
	Unimpaired	g	h	i	g+h+i
	Total	a+d+g	b+e+h	c+f+i	sum(a:i)

This portion of the analysis will be carried out on the primary endpoint population and will only include scores from the first examination.

2. Secondary Analyses of the Primary Endpoint

To better visualize the relationship between the two tests, the DCTclock and MMSE continuous scores will be plotted separately against the MoCA scores in a scatter plot, superimposed with a line of identity.

Linear and rank linear regression analyses will be conducted and p-values and 95% CI will be calculated for the y-intercept and slope of each model. Either the Pearson (linear regression) or the Spearman (rank regression) correlation coefficients will be produced, and superimposed onto the scatter plots for each regression method, along with 95% confidence intervals and the estimated regression line.

From Table I, the following statistics will be calculated for each of the tables (DCTclock vs MoCA and MMSE vs MoCA):

Positive Percent Agreement (PPA) = a/(a+d+g)*100 and its 95% Wilson CI

Negative Percent Agreement (NPA) = (e+f+h+i)/(b+c+e+f+h+i)*100 and its 95% Wilson CINo formal hypotheses will be carried out on the PPA and NPA calculations, although 95% CIs will accompany them.

5.1.2 Secondary Effectiveness Analysis (Test/Re-Test Reliability)

The secondary effectiveness analysis will assess the comparability of DCTclock test-retest reliability to that of MMSE. The secondary effectiveness analysis will be conducted on the evaluable population and will include two measurements, separated by 1-4 weeks, of each test of DCTclock and MMSE. Patients with "unanalyzable" DCTclock results will be excluded from the primary analysis of the secondary effectiveness analysis. Also, participants who no longer satisfy the inclusion/exclusion criteria prior to the second examination will also be excluded from this analysis. A scatterplot of test 1 vs test 2 with a superimposed line of identity will be created for both DCTclock scores and MMSE scores.

1. For each of the DCTclock and MMSE tests, a scatter plot of measurement 1 and measurement 2 will be generated and superimposed with a line of identity. An unweighted Deming regression will be carried out to determine the y-intercept, slope, and 95% CIs. The following hypothesis tests on the y-intercept and slope, respectively, will be carried out:

$H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$ and $H_0: \beta_1 = 1$ vs. $H_0: \beta_1 \neq 1$

where β_0 and β_1 are the y-intercept and slope, respectively, from the Deming unweighted regression.

Significant p-values ($p<0.05$) would indicate non-statistically significant test/re-test reliability, for a single test, although in order to compare the test/re-test reliability of DCTclock to that of MMSE, 95% CIs will be produced for the slope for the unweighted Deming regression model for each test. Overlapping CIs would indicate that there is no statistically significant difference in the test/re-test reliability between the DCTclock and MMSE tests.

Similarly, linear and rank linear regression models will be created for each test with reading 1 as the independent variable, and reading 2 as the dependent variable. Similar hypotheses as those for the Deming regression will be tested, and 95% CIs for the slopes will be calculated and compared among the two tests.

2. For each of DCTclock and MMSE separately, a 3x3 table will be constructed based on the results from two measurements, as follows:

Table III. 3x3 table for making calculations between Reading 1 and Reading 2

		Reading 2			
		Impaired	Indeterminate/Intermediate	Unimpaired	**Total**
Reading 1	Impaired	a	b	c	a+b+c
	Indeterminate/Intermediate	d	e	f	d+e+f
	Unimpaired	g	h	i	g+h+i
	Total	a+d+g	b+e+h	c+f+i	sum(a:i)

From the above table the following statistics will be calculated, for each test:

PPA = a/(a+d+g)*100 and its 95% Wilson CI

NPA = (e+f+h+i)/(b+c+e+f+h+i)*100 and its 95% Wilson CI

A quadratic weighted kappa score and its corresponding Wald 95% CI

Percent agreement for impaired = PPA = a/(a+d+g)*100 and its 95% Wilson CI

Percent agreement for indeterminate/intermediate = e/(b+e+h)*100 and its 95% Wilson CI

Percent agreement for unimpaired = i/(c+f+i)*100 and its 95% Wilson CI

Similarly, a 2x2 table for each test will be created, removing the indeterminate/intermediate results and the following statistics will be calculated:

PPA = a/(a+c)*100 and its 95% Wilson CI

NPA = d/(b+d)*100 and its 95% Wilson CI

For all of the above analyses in 2., 95% CI will be calculated and compared between DCTclock and MMSE. Overlapping CIs will be considered non-statistically significant.

3. A chi-square test will be used to assess the degree of association between the cognitive classifications of the two readings for each of DCTclock and MMSE. The null hypothesis (H_0) is that there is no association between the cognitive classifications of reading 2 and reading 1; the alternative hypothesis (H_1) is that there is an association between the classifications of the two readings. A p-value < 0.05 will be considered statistically significant and result in rejecting H_0 in favor of H_1.
4. If results show a significant difference between test and retest for DCTclock, that does not imply that DCTclock is less reliable than MMSE or other neuropsychological tests. Wald 95% CI will be calculated for DCTclock and MMSE for the quadratic weighted kappa between reading 1 and reading 2, separately. If the two confidence intervals overlap, this would suggest there is not a statistically significant difference between the test/re-test reliability of DCTclock and that of the MMSE. We will also look at the Wilson 95% CIs for PPA and NPA and compare them between DCTclock and the MMSE.

5.1.3 Additional Effectiveness Analyses

1. Construct Validity

To assess construct validity, the repeated administration of DCTclock scores will be compared to the scores from the repeated administration of a battery of neuropsychological tests. This will help to characterize the psychometric properties of DCTclock and to compare those properties to those found in MMSE.

After checking for normality, either Pearson or Spearman correlation coefficients will be used to assess correlation between the DCTclock scores and scores from the neuropsychological tests. This method will also be used to assess the correlation between the MMSE scores and scores from the neuropsychological tests.

Scatter plots of the DCTclock scores vs the scores of each separate neuropsychological test will be presented with a Deming regression line superimposed on the scatterplot.

All correlation coefficients for each test in the battery with DCTclock and MMSE will be formatted into a table with associated p-values and sample sizes.

2. Sensitivity Analyses

All above analyses involving MMSE scores may be repeated with different definitions of "Indeterminate/Intermediate" for MMSE. Given that the thresholds to determine impaired status from the MMSE range from <28 to <25 in clinical practice, the analyses will be repeated using various sets of thresholds to define MMSE unimpaired, indeterminate/intermediate and impaired status. This analysis is exploratory in nature and will only be used to explore the effect of different MMSE thresholds on the results.

3. Subgroup Analyses

All above effectiveness analyses (with the exception of the formal hypothesis test of the weighted kappa) may be repeated by (1) age (below median age, at or above median age); (2) sex; (3) education level; (4)

language fluency test score (below median, at or above median) (5) medical history; and (6) current medications.

In addition, it is expected that the DCTclock test's ability to identify impaired participants exceeds that of MMSE when participants are mildly impaired. Thus, for participants who score unimpaired on MMSE but impaired on DCTclock, scores on the neuropsychology battery test will be further inspected as follows to assess how often the DCTclock yielded the correct diagnosis in these cases. Specifically, the following will be presented on this subset of participants:

- Percentage of participants scoring in the impaired range on the MoCA test (25 or below)
- Percentage of participants scoring 1.5 standard deviations below the mean on two or more of the following tests (Benton Judgement of Line, Rey Complex Figure, DKEFs, Digit Span, Blocks, Trails, Symbol Digit).
- Percentage of participants 2 standard deviations below the mean on one of the following tests (Benton Judgement of Line, Rey Complex Figure, DKEFs, Digit Span, Blocks, Trails, Symbol Digit).

5.1.4 Safety Analysis

The frequencies and percentages of patients experiencing a serious/severe device-related adverse event (determined to be at least "probably related" to the device) will be summarized. This will be conducted on the Safety population.

5.1.5 Additional Summaries

All demographic and baseline characteristics will be summarized for both the ITT and Evaluable populations and descriptively compared between the two populations. Also to be inspected will be the baseline characteristics for the ITT participants not in the Evaluable population.

Lastly, demographic and baseline characteristics will be summarized stratified by DCTclock reading 1 results, and again by MMSE reading 1 results in order to compare these characteristics across identified groupings between the two tests. This table will also help drive the subgroup analyses.

For continuous and ordinal variables, descriptive statistics including the mean, standard deviation, median, minimum, maximum, interquartile range (IQR) and sample size will be included. For categorical variables, frequencies, percentages, and total sample sizes within groups will be presented.

5.2 Data Handling Procedures

5.2.1 Data Collection, Entry, and Quality Control

Data from the DCTclock test will be recorded electronically at the study site and sent to a secure, HIPAA-compliant server, while all other data will be recorded on paper and sent back to the sponsor site. Data entry for all hardcopy data will occur at the sponsor site and will be quality controlled (QC). QC of the data will occur at the sponsor site and individuals responsible for QC of the DCTclock test data will be distinct from individuals responsible for QC of the other neuropsychological tests' data.

5.2.2 Missing Data

There will be no imputation of missing data, and thus all analyses will be conducted using a listwise deletion (complete case analysis) approach. All sample sizes will be summarized for each analysis and the number of missing values for each variable will be tabulated.

5.2.3 Rounding

All p-values will be presented to three decimal places, while all other results will be presented to two decimal places.

5.3 Statistical Software

All data management, processing, statistical summarization, and statistical analyses will be performed using R (R Development Core Team, 2008; R Development Core Team, 2014) version 3.4.1 or higher on Windows. R, when used in a qualified fashion, can support the regulatory requirements for validated systems, thus ensuring that resulting electronic records are “trustworthy, reliable and generally equivalent to paper records."

6 References

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement, 20*(1), 37. doi:10.1177/001316446002000104

Fleiss, J. L., & Cohen, J. &. (1969). Large sample standard errors of kappa and weightred kappa. *Psychological Bulletin*, 323-327. doi:10.1037/h0028106

Halton, G. F. (1951). Note on an exact treatment of contingency goodness-of-fit and other problems of significance. *Biometrika, 38*, 141-149.

Linnet, K. (1993). Evaluation of regression procedures for methods comparison studies. *Clinical Chemistry, 39*(3), 424-432.

Linnet, K. (1998). Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clinical Chemistry, 44*(5), 1024-1031.

Nasreddine, Ziad S., et al (2005). "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment." *Journal of the American Geriatrics Society* 53.4 695-699.

Vanbelle, S., and A. Albert. “A bootstrap method for comparing correlated kappa coefficients.” *Journal of Statistical Computation and Simulation*, vol. 78, no. 11, 2008, pp. 1009–1015., doi:10.1080/00949650701410249.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* (R. F. Computing, Ed.) Retrieved from http://www.R-project.org

R Development Core Team (2014) R: Regulatory Compliance and Validation Issues A Guidance Document for the Use of R in Regulated Clinical Trial Environments. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/doc/R-FDA.pdf

[redacted]