Title: A Single Center Pilot Study to Evaluate Real Time Passive and Active High-Frequency Cognitive and Mood Assessment Data in Major Depressive Disorder Using Digital Wearable Technology

NCT Number: NCT03067506

SAP Approve Date: 13 July 2017

Certain information within this Statistical Analysis Plan has been redacted (ie, specific content is masked irreversibly from view with a black/blue bar) to protect either personally identifiable information or company confidential information.

This may include, but is not limited to, redaction of the following:

- Named persons or organizations associated with the study.
- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.
- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

Statistical Analysis Plan (SAP)

Takeda – MDD5003

Document Information

Type of Document :	SCI	ID / Version :	SCI-02.MDD5003v0.1
Developer :			
Template ID / Version :	TSCI-02 v1.2	Privacy :	Confidential

Approval by the following personnel signifies agreement that the content, rationale and approach are consistent with requirements, that a review for clarity, accuracy, completeness and compliance has been completed, and the document is approved for use.

Cognition Kit Approval

Name:	PPD	Role:	Principal Investigator
Signature:		Date:	13/07/17

Takeda Approval

Name:		Role:	
Signature:		Date:	

Approval by the Quality Manager (QA) indicates that a review has been completed for clarity, accuracy, completeness and compliance with company standards and regulatory requirements and that the document is approved for use.

QA Approval

Name:		Role:	Quality Analyst
Signature:		Date:	

Document History

Version	Date released for approval	Developer Initials	Reviewer Initials	Changes from Previous Version
1.0	19.June.2017	FC	PD	New document
1.1	11.Jul.2017	FC	PD	Included Draft Tables
1.2	13.Jul.2017	FC	PD	Minor edits for clarity

Contents

List of Abbreviations

CANTAB	Cambridge Neuropsychological Test Automated Battery
CI	Confidence Interval
ERT	Emotion Recognition Test
CCI	[redacted]
CCI	[redacted]
RVP	Rapid Visual Information Processing
SWM	Spatial Working Memory
CCI	[redacted]
LS	Least Squares
MMRM	Mixed Model Repeated Measures
SAP	Statistical Analysis Plan
SE	Standard Error
SD	Standard Deviation

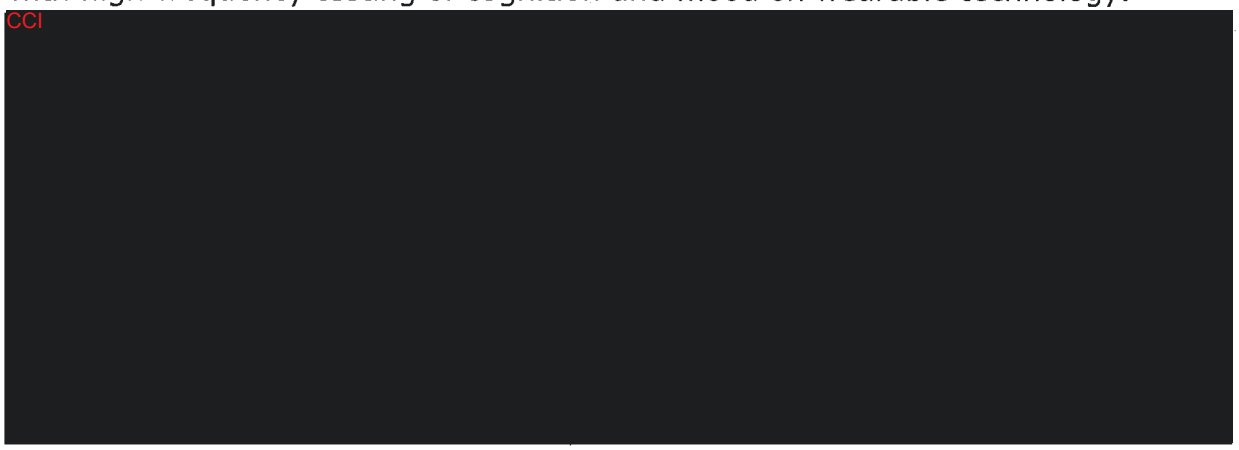
1. Introduction

This statistical analysis plan details the methodology and data analysis procedures that will be used for reporting the results analysis of data for the Takeda MDD-5003 study.

The study is a single-arm, unblinded, 6-week prospective observational feasibility study, designed to assess the feasibility and compliance with a novel method for assessing mood and cognition in subjects with major depressive disorder (MDD). The study assesses the feasibility of and compliance with cognitive and mood testing using wearable technology and the correlation of mood and cognition outcomes on wearable technology with traditional objective neuropsychological cognitive function tests and self-reported mood outcomes. Thirty subjects aged between 18 and 65 years, inclusive, with mild-moderate depression prescribed second- or third-line antidepressant monotherapy will be recruited. Subjects will be provided with an Apple Watch on which brief cognitive and mood tests will be administered daily. Subjects will take part in up to 5 study visits, 1 in-person on-site visit, 3 web-based and 1 at home visit assessing performance on traditional objective neuropsychological cognitive function tests and self-reported measures of depression symptom severity and social function.

The co-primary endpoints are to evaluate feasibility of and subject compliance with high-frequency testing of cognition and mood on wearable technology.

CCI



Study Design - Treatment Groups and Subjects

The sample consists of 30 subjects aged between 18 and 65 years. Subjects should have a diagnosis of depression, and scores on the PHQ-9 at screening which place them in the mild-moderate range for severity (PHQ-9 score between 5 and 15). Subjects are assessed on self-reported measures and CANTAB cognitive tests at four time points: Familiarisation, Baseline, Week 3 and Week 6. Subjects will be provided with an Apple Watch at the Baseline visit and will be prompted to complete three n-back tests three times each day and a questions assessing mood and cognitive symptoms.

As the primary endpoint is compliance with cognitive testing, the full cohort will be analysed. The CANTAB data will be analysed, excluding familiarisation. The familiarisation results will be reported separately for transparency.

2. Data to be analysed

2.1. Subjects: Analysis Population

All subjects who are assessed at baseline will be included in the analysis.

2.2. Outcome Measures

The variables to be analysed are shown in Table 1 with a full description available in Appendix A.

Table 1 Outcome variables

Outcome variable	Variable code
Daily n-back compliance	nBackN
Daily mood compliance	mood
Daily watch worn	watch
N-Back d'	dprime
N-Back Hit RT	hitRT
N-Back FA RT	FART
N-Back Hit Coefficient of Variation	hitCv

Daily depressed mood	DepressedMood
Daily self-reported cognitive problems	SelfReportCognitive
Daily anhedonia	ReducedPleasure
Daily step-count iPhone	TotalStepCountIPhone
Daily step-count Apple Watch	TotalStepCountWatch
Mean daily heart rate	MeanHeartRate
Minimum daily heart rate	MinHeartRate
Maximum daily heart rate	MaxHeartRate
Standard Deviation daily heart rate	SDHeartRate
CANTAB Emotion Recognition Unbiased Hit Rate Anger	ERTUHRA
CANTAB Emotion Recognition Unbiased Hit Rate Disgust	ERTUHRD
CANTAB Emotion Recognition Unbiased Hit Rate Fear	ERTUHRF
CANTAB Emotion Recognition Unbiased Hit Rate Happiness	ERTUHRH
CANTAB Emotion Recognition Unbiased Hit Rate Sadness	ERTUHRS
CANTAB Emotion Recognition Unbiased Hit Rate Surprise	ERTUHRSU
CANTAB Rapid Visual Processing A-Prime	RVPA
CANTAB Rapid Visual Processing Latency Standard Deviation	RVPLSD
CANTAB Rapid Visual Processing Latency Median	RVPMDL
CANTAB Spatial Working Memory Between Search Errors	SWMBE

CANTAB Spatial Working Memory Strategy	SWMS
CCI	CCI
CCI	CCI
CCI Total Score	CCI

3. Statistical Analysis

3.1. Descriptive Statistics

Compliance across the study will be transformed into a binary variable for mood and n-back, with days where participants completed testing coded as 1 and days with no data coded as 0. The percentage of days with data will be computed overall and by study week and tabulated.

For each subject, n-back performance across the study will be characterised by four summary variables – the mean, capturing global levels of performance, root mean square of successive differences (RMSSD), capturing variability in performance, and two measures extracted from mixed-effect modelling of the data: intercept and slope, representing initial performance and learning. These summary measures will be tabulated.

Daily mood assessment will be summarised across the study, capturing the overall level of mood across the study whole, and by study week. Means, standard deviation, RMSSD, median, number of subjects and minimum and maximum will be calculated and tabulated.

Activity data from the phone and the watch and heart rate from the watch will be aggregated by day and study week, and means, standard deviation, median, number of subjects and minimum and maximum will be calculated and tabulated.

For each of the CANTAB and PRO outcome variables, mean values, standard deviation, median, number of subjects, and minimum and maximum values will be tabulated by time point.

The output will also contain box plots/ line graphs of mean values (absolute scores) by time point, along with individual subject trajectories for each outcome variable.

3.2. Primary analysis: Compliance

Compliance with daily mood testing and daily n-back testing will be analysed separately. We will report the overall compliance (% completed testing sessions) over the six weeks of the study. The relationship between overall rates of compliance with baseline subject characteristics (CCI) will be explored through a series of bivariate correlations. These will be Pearson correlations or Spearman's rank correlation coefficient, depending on the observed distribution of the variables. As this is an exploratory study, no adjustment will be made for multiple comparisons. CCI CCI CCI . This analysis will be carried out using a series of Chi-square tests.

3.2.1.Logistic Mixed-model (LMM)

We expect compliance with testing to decline over the course of the study, therefore we will use a logistic regression mixed effect model to analyse the effect of time on compliance. Day will be used as fixed factor, and subject as a random effect. We plan on testing the impact of covariates on compliance, including baseline CCI Covariates which do not make a significant contribution will not be carried forward. Final model will be selected on the basis of Bayes Information Criterion.

3.2.1.1. Model specification:

<< variable>=subject + day + covariate + covariate * day.

Parameters will initially be analysed without transformation, but if the data suggest otherwise, an appropriate transformation (e.g. logarithmic, square root or reciprocal) will be applied. The analysis will include the calculation of effect sizes. The assumptions of the logistic regression model will be investigated by

examining the distribution of residuals and the pattern of residuals versus fitted values. Where non-normality or non-constant variance is observed, a transformation will be applied so that the assumptions are satisfied.

Transformations will be selected by applying different transformations iteratively (e.g. logarithmic, square root, reciprocal) followed by visual inspection of the cumulative normal plot, histogram of residuals and plot of residuals vs. fitted values. The most suitable transformation in terms of normality and constant variance will be selected. If no suitable transformation can be found, then non-linear models will be considered.

3.3. Co-primary analysis: Relationship between n-back and CANTAB data

For each of the CANTAB outcome measures, CCI [redacted], the correlation between summary n-back metrics described above (mean, RMSSD, intercept and slope) will be reported with the corresponding 95% Confidence Intervals and p value. Correlations between aggregate scores on the daily mood assessment (weekly and overall means and variability measures (SD and RMSSD)) with CCI [redacted] assessments will be reported with the corresponding 95% Confidence Intervals and p-value.

Pearson correlations or Spearman's rank correlation coefficient will be calculated, depending on the observed distribution of the variables. As this is an exploratory study, no correction for multiple comparisons will be made.

3.4. Secondary analyses

CCI [redacted]

3.5. Qualitative data analysis

The qualitative analysis will draw on data about compliance and motivation emerging from the semi-structured study entry interviews, study technical support queries and logs, and semi-structured end of study interviews. It will be conducted through an interpretative phenomenological analysis, and involve 2 stages:

1. Data preparation, which will include:
    - Familiarization with the data, reading notes, and/or listening to the audio dialogue to extract main themes and ideas.
    - Thematic framework development, identifying the key issues and concepts present in the data and creating themes both inductively, based on the data, and deductively, based on the research questions.
2. Interpretation stage, which will include defining the main concepts and mapping the ways in which different parts of the data are related to each other.

4.5.1. Presentation of qualitative findings

The outcome of the data analysis process described above will be a set of findings that relate to participant's motivation and compliance over the 6 weeks in the study. The findings will be organized in themes, and link experiences of the daily cognition and mood tasks with broader contextual factors that contribute to interest in taking part, and varying abilities to do so, over time.

Once the qualitative data has been analysed, the findings will be presented through the following formats:

- Narrative summary including anonymized quotes, key themes and findings.
- Individual case studies for certain participants representing key themes.
- Tabular overview of main qualitative data points.

5. Software and format of statistical results

5.5. Software

The statistical analysis will be carried out using RStudio 0.99.903 or higher. All descriptions, tables, figures and listings will be provided in a computerised document.

5.6. Format of results

All table and figure headings will be in the format 'Table x.x: Title', for example, 'Table 2.1: Descriptive Statistics - RTI Median Five Choice reaction time.' The headers in all tables will be centralised and the data right-justified.

6. List of Tables

Indicative dummy tables are provided.

6.5. Demographic data

Table 1.1 Demographic data

Variable		Outcome
Gender	Female	19
	Male	11
Age	N	30
	Mean	37.23
	SD	11.5
	Min	19
	Max	63
Screening PHQ 9	N	30
	Mean	9.13
	SD	4.1
	Min	5
	Max	15
Comorbid Anxiety	Yes	10
	%	33.3
Current Medication	Amitriptyline n	4
	%	13.33
	Citalopram n	9
	%	30
	Escitalopram n	1
	%	3.33
	Fluoxetine n	2
	%	6.67

	Mirtazapine n	2
	%	6.67
	Sertraline n	8
	%	26.67
	Trazodone n	1
	%	3.33
	Venlafaxine n	3
	%	10
Current Medication by Category	SARI n	1
	%	3.33
	SNRI n	5
	%	16.67
	SSRI n	20
	%	66.67
	TCA n	4
	%	13.33
Time on Current Medication (days)	N	30
	Mean	297.86
	Min	11
	Max	1030
Medication Switch	Yes	24
	%	80

6.6. Descriptive Statistics

Descriptive statistics for outcome variables (N, Mean, SD, median, minimum, maximum) by time point, including baseline and familiarisation. For daily n-back RMSSD, intercept and slope will also be reported. For daily mood, RMSSD will also be reported.

Table 2.1 Descriptive Statistics – Daily n-back compliance

	Overall	Week1	Week2	Week3	Week4	Week5	Week6
Mean	94.16	92.63	90.78	93.55	96.31	98.62	93.09
Median	90	89	92	91	96	97	91
SD	5.5	4.3	5.6	4.9	4.2	5.0	4.8
Minimum	80.00	30.00	34.00	33.00	30.00	44.00	23.00

Maximum	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Overall Rate	92.95	92.63	90.78	91.86	95.43	96.40	90.58

6.7. Correlation Matrices

6.8. Qualitative analysis

7. List of Figures

7.5. Box Plots

Box plots of outcome variables by time point including screening/familiarization

7.6. Heat-map plots

7.7. Individual Subject Trajectories

Subject trajectories by time point including screening/familiarization session.

7.8. Scatter plots

8. Listings

9. References

1. <<Protocol version and date>>
2. <<CANTAB Connect User Requirements Specification (URS) inc Date>>

Appendix A Cantab Outcome Measures

ERT	ERTOMDRT	KEY: ERT Overall Median Reaction Time:
		The overall median latency for a subject to select an emotion word after being presented with a stimulus. Calculated across all assessed trials.
ERT	ERTTH	KEY: ERT Total Hits:
		The total number of correct responses (emotion selection) the subject made across all assessed trials.
ERT	ERTUHRH	KEY: ERT Unbiased Hit Rate Happiness:
		The unbiased hit rate ensures that recognition accuracy of the Happiness emotion is not influenced by response guessing or response bias effects. It takes into consideration the joint probability of an individual making a correct response, based on the presentation of the correct stimulus out of the available possibilities. Calculated for assessed Happiness trials only.
ERT	ERTUHRS	KEY: ERT Unbiased Hit Rate Sadness:

		The unbiased hit rate ensures that recognition accuracy of the Sadness emotion is not influenced by response guessing or response bias effects. It takes into consideration the joint probability of an individual making a correct response, based on the presentation of the correct stimulus out of the available possibilities. Calculated for assessed Sadness trials only.
ERT	ERTUHRF	KEY: ERT Unbiased Hit Rate Fear:
		The unbiased hit rate ensures that recognition accuracy of the Fear emotion is not influenced by response guessing or response bias effects. It takes into consideration the joint probability of an individual making a correct response, based on the presentation of the correct stimulus out of the available possibilities. Calculated for assessed Fear trials only.
ERT	ERTUHRA	KEY: ERT Unbiased Hit Rate Anger:
		The unbiased hit rate ensures that recognition accuracy of the Anger emotion is not influenced by response guessing or response bias effects. It takes into consideration the joint probability of an individual making a correct response, based on the presentation of the correct stimulus out of the available possibilities. Calculated for assessed Anger trials only.
ERT	ERTUHRSU	KEY: ERT Unbiased Hit Rate Surprise:
		The unbiased hit rate ensures that recognition accuracy of the Surprise emotion is not influenced by response guessing or response bias effects. It takes into consideration the joint probability of an individual making a correct response, based on the presentation of the correct stimulus out of the available possibilities. Calculated for assessed Surprise trials only.

ERT	ERTUHRD	KEY: ERT Unbiased Hit Rate Disgust:
		The unbiased hit rate ensures that recognition accuracy of the Disgust emotion is not influenced by response guessing or response bias effects. It takes into consideration the joint probability of an individual making a correct response, based on the presentation of the correct stimulus out of the available possibilities. Calculated for assessed Disgust trials only.
RVP	RVPA	KEY: RVP A′
		A′ (A prime) is the signal detection measure of a subject's sensitivity to the target sequence (string of three numbers), regardless of response tendency (the expected range is 0.00 to 1.00; bad to good). In essence, this metric is a measure of how good the subject is at detecting target sequences.
RVP	RVPPH	RVP Probability of Hit:
		The number of target sequences during assessment blocks that were correctly responded to within the time allowed, divided by the number of target sequences during assessment blocks.
		(Correct hits ÷ total number of sequences)
RVP	RVPPFA	KEY: RVP Probability of False Alarm

		The number of sequence presentations that were false alarms divided by the number of sequence presentations that were false alarms plus the number of sequence presentations that were correct rejections
		(False Alarms ÷ (False Alarms + Correct Rejections))
RVP	RVPTH	RVP Total Hits
		The total number of target sequences that were correctly responded to (Correct Hits) within the allowed time during assessment sequence blocks.
RVP	RVPTFA	RVP Total False Alarms
		The total number of stimulus presentations during assessment blocks that were false alarms.
RVP	RVPTM	RVP Total Misses
		The total number of target sequences that were not responded to within the allowed time during assessment sequence blocks.
RVP	RVPML	RVP Mean Response Latency
		The mean response latency on trials where the subject responded correctly. Calculated across all assessed trials.
RVP	RVPMDL	KEY: RVP Median Response Latency

		The median response latency on trials where the subject responded correctly. Calculated across all assessed trials.
SWM	SWMBE	KEY: SWM Between Errors
		The number of times the subject incorrectly revisits a box in which a token has previously been found. Calculated across all assessed four, six and eight token trials.
SWM	SWMS	SWM
		The number of times a subject begins a new search pattern from the same box they started with previously. If they always begin a search from the same starting point we infer that the subject is employing a planned strategy for finding the tokens. Therefore a low score indicates high strategy use (1 = they always begin the search from the same box), a high score indicates that they are beginning their searches from many different boxes. Calculated across assessed trials with 6 tokens or more.