A three-arm, parallel group, multicentre, double-blind, randomized controlled trial evaluating the impact of GeneSight Psychotropic and Enhanced-GeneSight Psychotropic, on response to psychotropic treatment in outpatients suffering from a major depressive disorder (MDD) and having had – within the current episode - an inadequate response to at least one psychotropic medication included in GeneSight Psychotropic

NCT02466477

April 23rd, 2019

Statistical Analysis Plan

ARX 1009 GAPP MDD

Trial Full Title	**A three-arm, parallel group, multicentre, double-blind, randomized controlled trial evaluating the impact of GeneSight Psychotropic and Enhanced-GeneSight Psychotropic, on response to psychotropic treatment in outpatients suffering from a major depressive disorder (MDD) and having had – within the current episode - an inadequate response to at least one psychotropic medication included in GeneSight Psychotropic**
Trial Short Title	GAPP MDD Study
SAP Version	1.0
SAP Version Date	April 2, 2019
Trial Statistician	Paige Davies, M.Sc.
Trial Principal Investigators	Jim Kennedy, MD and Bryan Dechairo, Ph.D.
SAP Author	Paige Davies, M.Sc.
Sponsor Clinical Monitor	Ana Gugila

Issued by: Paige Davies _____ Date: 4/23/2019

Paige Davies, MS
Biostatistician I
Assurex Health, Ltd.

Approved by: J L Kennedy MD _____ Date: 04/23/19

James Kennedy, MD
Director, Neuroscience Research Department and Head, Psychiatric Neurogenetics Sections, CAMH
Professor, Department of Psychiatry and Institute of Medical Science, University of Toronto

Approved by: _____ Date: 4/5/19

Bryan Dechairo, PhD
Executive Vice President, Clinical Development
Myriad Genetics, Inc.

Tables of Contents

1 INTRODUCTION

1.1 Preface

Major depressive disorder (MDD) is a highly prevalent mental disorder and a leading source of disease burden worldwide with a prevalence of 8.7% in the United States (U.S.) and 8.1% in Canada (Vasilidis et al., 2007). The GeneSight Psychotropic product is a pharmacogenomic decision support tool that helps clinicians to make informed, evidence-based decisions about proper drug selection, based on the testing for clinically important genetic variants in multiple pharmacokinetic and pharmacodynamic genes that affect a patient's ability to tolerate or respond to medications.

This statistical analysis plan (SAP) is based on the study protocol EXCITE_AssureRx_CAMH_Protocol (MDD), version #1.5, dated on May 2, 2017.

It is expected that results from this trial will be used to inform guidelines for the use of pharmacogenomic testing in the treatment of MDD. Results may also be shared with regulatory bodies in Canada and abroad. In addition, the results from this analysis will be presented to Health Quality Ontario in regard to the Health Technology Application for GeneSight Psychotropic.

1.2 Objective of the Analyses

The analyses detailed in this document will assess the efficacy and safety of the patients treated with GeneSight guidance in comparison with the patients treated without GeneSight guidance. This analysis plan supersedes the one in the protocol and the previous versions.

2 STUDY OBJECTIVES AND DEFINITIONS

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective of this study is to evaluate the impact of GeneSight Psychotropic on response to psychotropic treatment as judged by the mean percentage change in the 17-item Hamilton Depression (HAM-D17) score from baseline to end of Week 8 of the study.

H_o: There is no difference in mean percentage change from baseline in HAM-D17 between GeneSight (GEN) and Treatment As Usual (TAU) at Week 8

H_a: The mean percentage change from baseline in HAM-D17 for GEN is different from TAU at Week 8

2.1.2 Secondary Objectives

The following secondary analyses will be performed in order of appearance.

2.1.2.1 Evaluate the following for GEN vs TAU:

- Percentage of responders in HAM-D17 at Week 8 in each treatment group (see definition in 7.1)
- Percentage of remitters in HAM-D17 at Week 8 in each treatment group (see definition in 7.2)

2.1.2.2 Evaluate the following for GEN vs E-GEN:

- Mean percentage change from baseline in HAM-D17 at Week 8
- Percentage of responders in HAM-D17 at Week 8
- Percentage of remitters in HAM-D17 at Week 8

The treatment groups used for all subsequent analyses (all scales at each time point) will depend on the results in 2.1.2.2

- If GEN is not statistically different than E-GEN at any endpoint, GEN and E-GEN treatment arms will be combined and compared to TAU
- If E-GEN is statistically superior to GEN at any endpoint, E-GEN will be compared to TAU
- If GEN is statistically superior to E-GEN at any endpoint, GEN will be compared to TAU

2.1.2.3 The subsequent secondary objectives are to evaluate the following for treatment groups determined in 2.1.2.2:

- Mean percentage change in HAM-D6 (see definition in 7.3) from baseline to Week 8;
- Mean percentage change in the 16-item Quick Inventory of Depression Symptomology (QIDS-SR16) from baseline to Week 8;
- Mean percentage change in the 9-item Patient Health Questionnaire (PHQ-9) from baseline to Week;
- Percentage of responders at Week 8 in each treatment group on the HAM-D6, QIDS-SR16, PHQ-9 (see definition in 7.1);
- Percentage of remitters at Week 8 in each treatment group on the HAM-D6, QIDS-SR16, and PHQ-9(see definition in 7.2)

2.1.3 Long-term Secondary Objectives

The long-term secondary objectives are to evaluate the following for treatment groups determined from results of 2.1.2.2:

- Mean percentage change in HAM-D17, HAM-D6, QIDS-SR16, PHQ-9, from baseline to Week 12, Month 6, Month 9, Month 12, (across all scales, where applicable) in each treatment group;
- Percentage of responders at Week 12, Month 6, Month 9, and Month 12 in each treatment group on the HAM-D17, HAM-D6, QIDS-SR16, PHQ-9, (across all scales, where applicable) (see definition in 7.1);
- Percentage of remitters at Week 12, month 6, month 9, and month 12 in each treatment group on the HAM-D17, HAM-D6, QIDS-SR16, PHQ-9 (across all scales, where applicable) (see definition in 7.2);

2.1.4 Exploratory, Safety, and Tolerability Objectives

- Mean percentage change in Generalized Anxiety Disorder 7-item (GAD-7) scale from baseline to end of Week 8, Week 12, Month 6, Month 9, Month 12;
- Percentage of responders and remitters at Week 8, Week 12, Month 6, Month 9, Month 12 on GAD-7 (see definition in 7.1 and 7.2);
- Change in SF-36 from baseline to end of Week 12, Month 6, Month 9, and Month 12;
- Percentage of responders at Week 12, Month 6, and Month 12 on SF-36 (see definition in 7.1);
- Mean percentage change in the Clinical Global Impression: Severity of Illness (CGI-S) from baseline to end of Week 12 and Month 12;
- Percentage of responders and remitters at Week 12 and Month 12 on CGI-S in each treatment group (see definition in 7.1 and 7.2);

- Percentage responders at Week 12 and Month 12 on CGI-I (Clinical Global Impression: Global Improvement) in each treatment group (see definition in 7.1);
- Percentage responders at Week 12 and Month 12 on CGI-IE (Clinical Global Impression: Efficacy Index) in each treatment group (see definition in 7.1);
- Compare congruence of medications between treatment groups at Week 8, Week 12, Month 6, Month 9, and Month 12 (see 7.5 for definition);
- Percentage of subjects who experienced adverse events (AE) at Week 8, Week 12, and Month 12;
- Change in Udvalg for Kliniske Undersogeler (UKU) side effect rating scale; Frequency, Intensity and Burden of Side Effects Ratings (FIBSER); and weight gain from baseline to Week 8, Week 12, and Month 12;
- Percentage of patients compliant with medications at Week 8, Week 12, Month 6, Month 9, and Month 12 between treatment groups using Brief Adherence Rating Scale (BARS) and prescription filling data

2.2 Endpoints

2.2.1 Primary Efficacy Variable:

- Percentage change in HAM-D17

2.2.2 Secondary Efficacy Variables:

- Percentage change in HAM-D6
- Percentage change in QIDS-SR16
- Percentage change in PHQ-9
- Responders of HAM-D17, HAM-D6, QIDS-SR16, and PHQ-9
- Remitters of HAM-D17, HAM-D6, QIDS-SR16, and PHQ-9

2.2.3 Exploratory, Safety, and Tolerability Variables

- Percentage change in GAD-7
- Percentage change in CGI-S
- Change in SF-36
- Responders of GAD-7, CGI-S, CGI-I, CGI-EI, and SF-36
- Remitters of GAD-7, and CGI-S
- Change in EuroQol (EQ-5D-5L)
- Change in BARS
- Percentage of patients on congruent medications
- Prescription filling frequency between study visits
- Number of Adverse Events (AE)
- Change in UKU

- Change in FIBSER

3 GENERAL CONSIDERATIONS

3.1 Timing of Analyses

A definitive statistical analysis of the primary and 8-week secondary outcome measure will be performed by the trial statistician when all of the following have been achieved:

- All subjects recruited into the study have been followed up for 8 weeks or have been deemed to be lost to follow-up;
- All CRFs have been entered onto the computer database;
- All data have been checked for completeness, and the accuracy of all data entries have been verified; and
- Evaluability status of subjects has been determined;

Long-term and secondary outcome measures of efficacy and safety will be performed by the trial statistician when all of the following have been achieved:

- All subjects recruited into the study have been followed up for 12 months or have been deemed to be lost to follow-up;
- All CRFs have been entered onto the computer database;
- All data have been checked for completeness, and the accuracy of all data entries have been verified;
- Evaluability status of subjects has been determined; and
- Database has been locked.

3.2 Analysis Populations

3.2.1 Per Protocol (PP) Population

The primary statistical analysis will be performed using the Per Protocol (PP) principle. The population of this analysis will be all subjects who met the inclusion and exclusion criteria. Other variables relating to exclusion may include but are not limited to:

At screening:

a) Meets all inclusion/exclusion criteria (except exception 8 and 13)
b) All assessments completed by appropriate person
c) QIDS-C16 and -SR16 are $\geq$ 11

At baseline:

a) Confirmation that subject still satisfies all Inclusion/Exclusion criteria
b) Randomization occurred prior to the Baseline visit

c) All assessments completed by appropriate person. The Blinded Rater should administer the QIDS-C16, not the Treating Clinician.
d) Baseline was completed within 42 days of screening
e) Sample was received and report was released (GEN & E-GEN) prior to baseline
f) Report was viewed (GEN & E-GEN) at/prior to baseline
    - If report not viewed at/prior to baseline, prescription decision date must follow view date of report
g) Baseline QIDS-SR is $\geq$ 11
h) Baseline HAMD-17 $\geq$ 14

After enrollment:

a) Week 4 visit occurs within 20-36 days of baseline visit
b) Week 8 visit occurs within 48-64 days of baseline visit
c) QIDS-C16 completed by Blinded Rater, not Treating Clinician
d) Subject did not start ECT, DBS, or TMS
e) MDD Rx given by study affiliated clinician
f) If unblinding occurs before week 8 visit, data is not included in PP population

3.2.2 Intention to Treat (ITT) Population

The population for this analysis will be all subjects who met the inclusion and exclusion criteria at screening, and were randomized. Subjects that meet inclusion and exclusion criteria and were randomized but have low baseline QIDS-SR scores will be included in ITT analysis.

3.3 Covariates

Baseline HAM-D17 score is a covariate for the primary analysis. For secondary and exploratory analyses, the following covariates may be added to the model with baseline score: Age, Gender, Treatment, and Site Type (PCP/Psych).

3.4 Missing Data

Missing observations of response variables will be checked and missing patterns by treatment and overall will be described including a CONSORT trial flow diagram. Missing values will be handled by using maximum likelihood method (ML). Complete cases analyses will be conducted and results will be compared to those obtained from ML. The main analysis conclusion at Week 8 will be drawn using ML method for the PP population (see Section 5.1).

4 SUMMARY OF STUDY DATA

All continuous variables will be summarized using the following descriptive statistics: (non-missing) sample size (n), mean or median, standard deviation or range (maximum and minimum).

All categorical variables will be summarized using the frequency and percentage (based on the non-missing sample size) for each observed category.

All summary tables will be structured with a column showing the appropriate summary statistics for all participants combined, separate columns for Treatment As Usual (TAU), GeneSight (GEN), and Enhanced GeneSight (E-GEN) groups; sample sizes and/or numbers of missing observations will also be reported. A final column will show, where appropriate, estimates of effect size with their 95% confidence intervals.

Separate tables will be provided for the PP and ITT analyses.

4.1 Demographic and Baseline Variables

Summary statistics for demographic and baseline variables will be produced and group differences will be computed for these variables.

5 EFFICACY ANALYSIS

5.1 Primary Efficacy Analysis

The primary measure of efficacy is the percentage change from baseline to week 8 in HAM-D17. The percentage change from baseline in HAM-D17 will be analyzed using a Mixed Model for Repeated Measures (MMRM). The model will include treatment, week (4 & 8), treatment-by-week interaction, baseline HAM-D17 score, baseline HAM-D17-by-week interaction as fixed effects. Unstructured covariance between measurements at weeks four and eight, from the same patient, will be incorporated into the model. If there is a convergence issue, Toeplitz covariance will be used. The MMRM method employed here is known as Maximum Likelihood (ML) to effectively handle the missing values as discussed in Section 3.4. The p-value will be derived from the T-test for comparing two treatment arms at week 8 (treatment by week interaction term). If overwhelming evidence suggests that the normality assumption is not satisfied, complete case analyses (CCA) of the primary efficacy endpoint will be conducted for the PP population for the completers of week eight. The primary endpoint will be analyzed by fitting an analysis of covariance (ANCOVA) model which includes baseline HAMD-17 score. Robust regression method (M Estimation Method by Huber) will be used for CCA to detect potential outliers and appropriately weigh the influence of the outliers should they exist.

As a sensitivity analysis, complete case analyses (CCA) of the primary efficacy endpoint will be conducted for the PP population for the completers of Week 8. The primary endpoint will be analyzed by fitting an analysis of covariance (ANCOVA) model which includes treatment, baseline HAMD-17 score, age, and gender. If age and gender are not statistically significant, they may be dropped from the final model. Robust regression method (M Estimation Method by Huber) will be used for CCA to detect potential outliers and appropriately weigh the influence of the outliers should they exist.

The above MMRM and CCA analyses will also be conducted for the ITT population.

The primary methods for checking the normality assumption will be through graphical assessment using a Q-Q plot and histogram of the residuals. If the points for the Q-Q plot are approximately linear and if the histogram is approximately bell-shaped, the residuals will be considered normally distributed.

The constant variance assumption will be checked by a scatter plot of residuals vs. predicted mean of response variable.

5.2 Secondary & Exploratory Efficacy Analyses

The secondary analyses will be conducted using MMRM method for continuous response variables and ANCOVA with robust regression for complete cases for the PP population and may also be conducted for the ITT population.

As a secondary analysis, the following covariates will be added to the primary endpoint model with baseline score and analyzed the same way: treatment, age, gender, site type, and race included in the model. If covariates are not statistically significant, they may be dropped from the final model.

The secondary and exploratory measures of efficacy including percentage change from baseline in HAM-D6, QIDS-SR16, PHQ-9, GAD-7, CGI-S, change in EuroQol (EQ-5D-5L), and change in SF-36 will be analyzed the same way as for the primary efficacy variable above with baseline score included in the model. The secondary measures of efficacy will also be analyzed with baseline score, treatment, gender, site type, age, and race included in the model. If covariates are not statistically significant, they may be dropped from the final model.

Percentage of responders for HAM-D17, HAM-D6, QIDS-SR16, PHQ-9, GAD-7, CGI-S, CGI-I, CGI-EI, and SF-36, and remitters for HAM-D17, HAM-D6, QIDS-SR16, PHQ-9, GAD-7, and CGI-S will be analyzed by visit, separately, using a Generalized Linear Mixed model. Pre-specified covariates may be added.

Percentage of patients on congruent and incongruent medications will be compared between GEN and TAU using Chi-square tests at each time point.

Medication compliance using BARS and prescription filling data will be summarized and compared between treatment groups (see 7.6 for description).

5.3 Safety and Tolerability Analysis

Descriptive statistics for the number of AE's by treatment and week will be generated. The percentage of subjects who experienced AE's may be analyzed using a Generalized Linear Mixed model. The model would include treatment, gender, age, and baseline HAM-D17 score. Continuous measures including UKU, FIBSER, and weight gain may be analyzed in a similar manner as for the primary efficacy variable above.

6 SIGNIFICANCE LEVEL

A significance level of 0.05 (2-sided) will be used. All analyses will be performed using IBM SPSS Statistics 25 and/or SAS 9.4 and/or JMP 14.

7 APPENDIX

7.1 Definition of Responders

For HAM-D17, HAM-D6, QIDS-SR6, HAMD-6, PHQ-9, GAD-7, SF-36, a responder is defined as a participant with 50% decrease from baseline in total scale score.

For Clinical Global Impression: Severity of Illness CGI-S: a responder is defined as a decrease in category of severity of at least 1 point, for Clinical Global Impression: Global Improvement CGI-I: it's defined as a score from 1 to 3, and for Clinical Global Impression: Efficacy Index CGI-EI: it's defined as scores of 01, 02, 05, or 06.

7.2 Definition of Remitters

A remitter is defined as a participant at a post-treatment visit with HAM-D17 ≤7, HAM-D6 <5[2], QIDS-SR6 ≤5, PHQ-9 <5, GAD-7 <5, or CGI-S ≤1.

7.3 Definition of HAM-D6

HAM-D6 is a subset of HAM-D17 and is the sum of score for the following questions: Depressed Mood (question 1), Feelings of guilt (question 2), Work and interests (question 7), Retardation (question 8), Anxiety – psychic (question 10), Somatic symptoms - general (question 13)[2].

7.4 Definition of Congruence

Congruence relates to whether the physician follows the combinatorial pharmacogenetic test recommendations. A three-level indicator variable (eg., green, yellow, and red categories) will be created using a proprietary algorithm which combines the phenotypes ascribed to each gene for each participant and drug metabolism information for each of the GeneSight panel drugs. Medications were considered congruent with the combinatorial pharmacogenomic test results if they were classified in green ('use as directed') or yellow ('use with caution') report categories. Incongruent medications were classified as those in red ('use with increased caution and more frequent monitoring') report category. Prescribing is considered congruent if a patient is prescribed only congruent medications. Prescribing was considered incongruent if a patient is prescribed one or more incongruent medications. Patients will be categorized as congruent or incongruent at baseline using prescribed medications reported at baseline and at Week 8 using prescribed medications reported at Week 8.

7.5 Brief Adherence Rating Scale (BARS)

The BARS is a recently developed clinician-administered adherence assessment tool consisting of three questions (adapted with permission from a questionnaire used in the CATIE trial) about the patient's knowledge of their own medication regimen and episodes of missed medication taking, as follows:

1) What is the total number of prescribed doses of medication you take per day?
2) What is the total number of days in the past month when you did not take the prescribed dose?
3) What is the total number of days in the past month when you took less than the prescribed dose?

8 REFERENCES

1. Vasiliadis, H.M., Lesage, A., Adair, C., Wang, P.S., Kessler, R.C., 2007. Do Canada and the United States differ in prevalence of depression and utilization of services? Psychiatr Serv 58(1), 63-71.
2. Kyle P.R., et al. 2016 Validity of the Different Versions of the Hamilton Depression Scale in Separating Remission Rates of Placebo and Antidepressants in Clinical Trials of Major Depression. J Clin Psychopharmacol (36), 453–456.