PROTOCOL TITLE: A 12-Week, Randomized, Double-Blind, Controlled Evaluation Followed by an Open-Label 12-Week Follow-up Period of the Impact of 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

NCT NUMBER: NCT02109939

DOCUMENT DATE: 23APRIL2015

STATISTICAL ANALYSIS PLAN

Trial Full Title	A 12-Week, Randomized, Double-Blind, Controlled Evaluation Followed by an Open-Label 12-Week Follow-up Period of the Impact of 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	TRD RCT
Protocol #	ARX1006
SAP Version	1.4
SAP Version Date	August 17, 2016
Trial Statistician	James Li, M.S., M.A.
Trial Principal Investigators	John Greden, MD and Sagar Parikh, MD
SAP Author	James Li, M.S., M.A.
Sponsor Clinical Monitor	Bryan Dechairo, Ph.D.

SIGNATURE PAGE

Approved by:	Date:	Signature:
James Li, MS, MA *Sr. Director of Statistics* Assurex Health, Inc.	8/31/16	
Bryan Dechairo, PhD *SVP, Medical Affairs and Clinical Development* Assurex Health, Inc.	8/31/16	
John F. Greden, MD *Principal Investigator – Trial* Executive Director, University of Michigan Comprehensive Depression Center Founding Chair, National Network of Depression Centers (NNDC) Chair Emeritus, University of Michigan Department of Psychiatry	8/30/16	
Sagar V. Parikh, MD, FRCPC *Principal Investigator – University of Michigan* Professor of Psychiatry, University of Michigan Medical School Associate Director, University of Michigan Comprehensive Depression Center	8/30/16	

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. Epidemiological studies estimate 12-month and lifetime prevalence for MDD in the United States to be 5.3% and 13.2%, respectively (reviewed in 2010). 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 trial is to replicate previous findings of improvement in clinical outcomes in subjects treated with GeneSight guidance vs. without GeneSight guidance in a double-blind, randomized control trial (RCT).

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

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 (HAMD-17) score from baseline to end of Week 8 of the study.

H_o: There is no difference in mean percentage change from baseline in HAMD-17 between GeneSight (GS) and Treatment As Usual (TAU) at Week 8

H_a: The mean percentage change from baseline in HAMD-17 for GS is different from TAU at Week 8

The secondary objectives are to evaluate the following:

- Mean percentage change in the 16-item Quick Inventory of Depression Symptomology (QIDS-C16) scale to end of Week 8 of the study;
- Mean percentage change in the 9-item Patient Health Questionnaire (PHQ-9) from baseline to end of Week 8 of the study;
- Mean percentage change in Clinical Global Impression of Severity (CGI-S) from baseline to end of Week 8;
- Clinical Global Impression of Improvement (CGI-I) from baseline to end of Week 8;
- Mean percentage change in HAMD-17 , QIDS-C16, PHQ-9, or Clinical Global Impression of Severity (CGI-S), and Clinical Global Impression of Improvement (CGI-I) scale from baseline to end of Week 12 of the study;
- Percentage of responders at Week 8 in each treatment group on the HAMD-17 , QIDS-C16, PHQ-9, CGI-S, CGI-I, or CGI-EI (see definitions in 7.1);
- Percentage of responders at Week 12 in each treatment group on the HAMD-17 , QIDS-C16, PHQ-9, CGI-S, CGI-I, or CGI-EI (see 7.1);
- Percentage of remitters at Weeks 8 and 12 defined as HAMD-17 $\leq$7, QIDS-C16 $\leq$5, PHQ-9 <5, CGI-S $\leq$1 in each treatment group (see 7.2);
- Mean percentage change in symptoms, percentage of response and percentage of remission from baseline to week 24 and from week 12 to week 24 across all scales in each treatment group.

The exploratory objectives of this study are to evaluate:

- The mean percentage change in Generalized Anxiety Disorder 7-item (GAD-7) scale from baseline to week 8 and week 12, and from week 12 to week 24 in each treatment group.

2.2 Endpoints

2.2.1 Primary Efficacy Variable:

- Percentage Change in HAMD-17

2.2.2 Secondary Efficacy Variables:

- Percentage change in QIDS-C16
- Percentage **c**hange in PHQ-9
- Percentage change in CGI-S
- Change in SF-36
- Percentage of responders of HAMD-17, QIDS-C16, PHQ-9, CGI-S, CGI-I, CGI-EI, & SF-36
- Percentage of remitters of HAMD-17, QIDS-C16, PHQ-9, & CGI-S

2.2.3 Exploratory Variables

- Percentage Change in GAD-7

2.2.4 Safety Variable

- Number of Adverse Events (AE)

3 GENERAL CONSIDERATIONS

3.1 Timing of Analyses

A definitive statistical analysis of the primary 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 24 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;
- Evaluability status of subjects has been determined; and
- Database has been locked.

3.2 Analysis Populations

3.2.1 Intention to Treat (ITT) Population

The population for this analysis will be all subjects who met the inclusion and exclusion criteria and were randomized.

3.2.2 Per Protocol (PP) Population

The population of this analysis will be all subjects who met the inclusion and exclusion criteria and followed the study protocol with moderate or higher depression severity at baseline - HAMD-17 score of 14 or above. The details of the PP population is defined in a separate document. The primary statistical analysis will be performed using the Per Protocol (PP) principle.

3.3 Covariates and Subgroups

Baseline HAMD-17 score is a covariate for the primary analysis. For secondary analyses, the following covariates will be added to the model with baseline score: Age, Gender, Site Type (academic/community), and Race. For exploratory analyses, the following covariates will be added to the model: Age, Gender, Site Size (Small & Large), and Race.

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) and GeneSight (GS) 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. No group difference statistics 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 HAMD-17. The percentage change from baseline in HAMD-17 will be analyzed using a Mixed Model for Repeated Measures (MMRM). The model will include treatment, week (4 & 8), treatment-by-week interaction, baseline HAMD-17 score, baseline HAMD-17-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 main conclusion regarding the primary endpoint will be drawn from this model on the PP population. 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 treatment and 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 and 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.

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 ITT population.

The secondary measures of efficacy including percentage change from baseline in QIDS-C16, PHQ-9, CGI-S, GAD-7, and change in SF-36 will be analyzed the same way as for the primary efficacy variable above.

Percentage of responders for HAMD-17, QIDS-C16, PHQ-9, CGI-S, CGI-I, and CGI-EI, and remitters for HAMD-17, QIDS-C16, PHQ-9, and CGI-S will be analyzed by visit, separately, using a Generalized Linear Mixed model. Pre-specified covariates may be added.

5.3 Safety 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 and baseline HAMD-17 score.

6 SIGNIFICANCE LEVEL

A significance level of 0.05 (2-sided) will be used. All analyses will be performed using SAS 9.4 and/or JMP 13.

7 APPENDIX

7.1 Definition of Responders

For HAMD-17, QIDS-C16, and PHQ-9, a responder is defined as a participant with 50% change from baseline in total scale score.

For Clinical Global Impression: Severity of Illness CGI-S: a responder is defined as a change 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 HAMD-17 ≤7, QIDS-C16 ≤5, PHQ-9 <5, or CGI-S ≤1.

Sample Size

In a previous smaller single site randomized control trail of GeneSight (intervention) an effect size of 0.3 in percent symptom improvement favoring the intervention group was observed with wide confidence intervals. The true effect size could be smaller in a larger multisite study due to factors including demographic heterogeneity, lower percent of subjects taking medications with gene-drug interactions at baseline and lower congruence with test recommendations. A total of 1,400 subjects will be enrolled for 1,200 completers giving 90% power at 5% level of significance (2-sided) to detect an effect size of approximately 0.2 favoring the intervention group in percent symptom change from baseline.

6960 Cintas Blvd.
Mason, OH 45040

513.701.5000
www.genesight.com