Heart Watch Study: a Pragmatic Randomized Controlled Trial

NCT04468321

Document Date: August 11, 2022

STATISTICAL ANALYSIS PLAN (SAP)

1.1 OVERALL STATISTICAL APPROACH AND ANALYSIS POPULATION

All analyses will be intention-to-treat. For the primary outcome, we will calculate the difference in the AFEQT global questionnaire score at baseline and at 6 months for patients randomized to the Apple Watch and patients randomized to the Withings Move arm.

The secondary clinical utilization outcomes will also be compared between the two groups using an intent-to-treat approach. Hugo will be the primary mechanism used to obtain clinical outcomes and clinical utilization, with comparisons between groups. We will also trend patients' monthly responses about anticoagulation adherence.

For the secondary outcomes for accuracy of Apple Watch Software Features that will be compared are a comparison of the heart rhythm (identical to 12-lead ECG or different), rate (difference in beats per minute), and PR, QRS, and QT intervals (difference in milliseconds) between what is assessed by the Apple Watch ECG feature and 12-lead ECGs during hospitalization or outpatient follow-up care. We will also compare the SpO2 readings taken by the Apple Watch at baseline to the measure performed by a standard, medical grade pulse oximeter.

All p-values will be significant at <0.05 with two-sided inferential tests. Continuous data will be presented as mean with standard deviation and categorical data as median with interquartile range.

1.2 BASELINE DESCRIPTIVE STATISTICS ANALYSIS

Baseline descriptive statistics will be reported for the overall study, site specific, and for both the control and treatment arms of the study. These will include patient age, gender race, ethnicity, and multiple comorbidities. We will compare baseline data by Chi Square (or Fischer's Exact) test for dichotomous/categorical variables and t-tests for continuous variables (unless the variables are non-parametric – in which case we will use a median test).

1.3 STUDY SUCCESS AND FAILURE CRITERIA

This study will be a success upon enrollment of 150 study participants and completion of 6 months follow-up.

1.4 SAMPLING PROCEDURE, SAMPLE SIZE AND POWER CALCULATION

The study sample size was calculated assuming 80% power to detect an effect size of 8.8 on the AFEQT questionnaire (slightly higher than the minimal clinically important difference of 5), with alpha 0.05.

We expect that drop-out rate will be close to zero in this study as patients are followed passively and the burden on patient is minimal. In terms of non-completion of the 6-month AFEQT Questionnaire, our

current study accounts for an estimated drop-out rate of 8% at 6 months based on the use of the AFEQT score from Holmes DN et al Circ Cardiovasc Qual Outcomes 2019, in which the mean ± SD was 79.0±18.5 and assuming an effect size of 8.8 on the AFEQT questionnaire, which is higher than the minimal clinically important difference of 5.

If patient drop-out or loss to follow-up occurs, then we will carry forward the last patient-reported outcome measure response. If an electronic health record data connection is lost for a study participant, we will only include the follow-up duration when data were available in the utilization endpoint analyses.

1.5 ACCEPTANCE CRITERIA

We will accept all data received from study participants. If patients are missing primary outcome data, we will use the last observation carried forward for the PROMs.

1.6 ENDPOINT ANALYSIS

1.6.1 PRIMARY ENDPOINT ANALYSIS

For the primary outcome, we will calculate the difference in the AFEQT global questionnaire score (as this is a continuous variable) at baseline and at 6 months and perform a comparison between the patients randomized to the Apple Watch and patients randomized to Withings Move. This will be calculated as per standard AFEQT calculations, which result in a summary score for this continuous outcome (which ranges from 0 to 100). We will use a t-test for this comparison. As this is an RCT, we expect that confounding will be minimal. If patients are missing outcome data, we will use the last observation carried forward for the patient-reported outcomes. Missing covariates will be set to missing.

At 12 months, we will also perform the same analysis, but this will be as an exploratory secondary endpoint. If a patient drops out, we will carry forward the most recent patient-reported outcome measure response.

We will also retrospectively compare AFEQT global questionnaire at months 1, 2, 3, 4, 5.

1.6.2 SECONDARY ENDPOINT ANALYSIS

The secondary clinical utilization outcomes will be compared between the two groups. Data obtained through Hugo will be used to obtain most outcomes. As these will intended to be composite outcomes, we will use the t-test or the Wilcoxon rank-sum test for comparison between the Apple Watch and Withings arms.

We will also use the t-test for comparison of the individual domains of the AFEQT questionnaire.

Electronic health records will be used to review ECG tracings. We will also trend patients' monthly responses about anticoagulation adherence. These analyses will be conducted for the entire studied population.

The secondary outcomes for accuracy of Apple Watch Software Features that will be compared are a comparison of the heart rhythm (identical to 12-lead ECG or different), rate (difference in beats per minute), and intervals: PR, QRS, and QT interval (difference in milliseconds) between what is assessed by the Apple Watch ECG feature and 12-lead ECGs during hospitalization or outpatient follow-up. Inter-observer and intra-observer agreement in assessment of the Apple Watch ECGs will be assessed using intraclass correlation coefficient (ICC).