Statistical Plan

Smartwatch Monitoring for Atrial Fibrillation After Stroke

NCT #: NCT03761394

Date of document: 11/20/18

Pulsewatch Statistical Plan:

Pulsewatch usability assessment via qualitative interviews and focus groups: Interviews will seek individual opinions about the usability and acceptability of Pulsewatch and will ask about: 1) ways in which Pulsewatch was intrusive; 2) how using Pulsewatch caused fatigue; 3) ways in which participants thought Pulsewatch use promoted health; 4) reflections on the trustworthiness of Pulsewatch (e.g., 'belief' of findings); and 5) participant comfort using Pulsewatch to communicate with their clinician. We anticipate that these 5 areas are highly individual and vary from person-to-person. Therefore, qualitative interviews, rather than standardized questionnaires, are needed since they provide an opportunity for users to describe their personal experiences in detail. In addition, we will ask participants to identify the amount of time they spent using Pulsewatch during a typical day. To assess another measure of participant's acceptance, we will ask the participant to gauge their interest in continuing to use Pulsewatch even though their participation in the study has concluded. Participants who ceased using Pulsewatch will be identified and contacted by the staff for semi-structured interviews to understand why they stopped. We will conduct 2-4 focus groups (12-16 users total) to generate data on barriers to, and facilitators of, Pulsewatch use. In distinction to Aim 1 focus groups, which focus on learning what capabilities users want, Aim 3 focus groups center on 'what worked' and 'what didn't' and whether or not long-term use introduces or reduces stress. We will also solicit feedback on the Interactivity Manager and Annotation Panels, features developed too late in Aim 1 to be fully explored by Aim 1 focus groups. Groups will be stratified by Pulsewatch adherence during the intervention period (<50%, 50-74%, >75%). We anticipate that with 4 focus groups we will reach saturation. However, if this is not the case, we will conduct 2 additional groups (n=12). Semi-structured and open-ended agendas will be written and IRB submitted. Discussions will be audio-recorded, transcribed and de-identified, protecting confidentiality by replacing names with study IDs.

Data Analysis: Sample Size and Adequacy of the Proposed Population. For Aim 1, we will recruit 30-40 stroke/TIA patients and 5-10 providers to participate in focus groups and interviews. 4 UMass Memorial Medical Center (UMMC) cardiologist/neurologists see 500 outpatients yearly with stroke/TIA in the UMMC neurology/cardiology clinic, requiring a recruitment rate of 3% over 6 months. We will recruit 120 participants with cryptogenic stroke admitted to the UMMC Stroke Service over a 33-month period. Since the UMMC Stroke Service discharges 143 patients with cryptogenic stroke yearly, we require a recruitment rate of 31% over 33 months.

Qualitative data analysis (Aims 1 and 3): Table 1 summarizes the planned qualitative analyses and outcomes.

Table 1. Qualitative Research Steps, Methods and Analyses

Research need	Aim	Method	Participants	Analysis	Outcome	Timing
Formative patient & provider input into Pulsewatch design & development	**1**	4-6 focus groups; 6-8 participants (ppts) each	30-40 app naïve ppts; 6-10 stroke medical providers will use a fully realized demo app and provide feedback on interfaces and usability	coding & thematic analysis	Feedback on prototype. Barriers & facilitators of use. Preferences & usability of features. Enhancements to text and ambient messaging. Provider information. Communication system design.	Before app refined
Communication among development team	**1**	Hack-a-thon	Patients, medical providers, behavioral scientists, computer programmers	coding & thematic analysis	Ensures developers hear patient and provider needs. Feedback on information display, graphical images, & message content.	App actively refined
Post-deployment eval of Pulsewatch	**3**	2-4 focus groups; 6-8 ppts each	12-16 ppts	coding & thematic analysis	Feedback on information display, graphical images & message content.	After app use
Post-deployment eval of Pulsewatch	**3**	20 min qualitative interviews	30 stroke survivors who use PULSEWATCH for 1 month	framework analysis	Impact of app use on quality of life, anxiety, & selfactivation	After app use

Qualitative analyses will involve a multi-step process, using applied thematic content analysis of all focus groups. In thematic analysis, all relevant qualitative data is assigned to specific codes (e.g., interface). The codes are then read in aggregate to identify key themes (e.g., large fonts needed). The Principle Investigator and co-investigator will develop transcript codes. The focus groups conducted in Aim 1 will be analyzed in Year 1 and will inform the Hack-a-thon; those conducted in Aim 3 will be analyzed separately in Year 4. Deductive codes will be drawn from the topics in the questions used to facilitate the focus groups and interviews. This coding scheme will be developed during close review of the transcripts, which also allows the creation of inductive codes to capture any themes that emerge from the discussions themselves. Two

coders will independently code each transcript. Coded transcripts will then be compared to ensure comprehensiveness and to reconcile discrepant results. Agreed upon codes will be entered into NVivo 10 (Version 10, QSR International, Australia) for analysis. NVivo 10 allows sorting of data based on relevant demographic variables. Thus, our thematic analysis will investigate whether there are differences in content related to sex or whether pAF was diagnosed. A framework analysis will be conducted on the 30 semi-structured summative interviews with Pulsewatch users (Aim 3). This qualitative data reduction technique is used with large data sets to review, summarize and classify data; it is particularly appropriate for practice-oriented findings and it is often used in health-related research.[113] NVivo 10 includes a framework analysis tool and will be used to manage data and facilitate the analyses.

Quantitative data analysis (Aims 2 and 3): Table 2 summarizes the primary quantitative analyses and outcomes. **As for Aim 1, for all Aim 2 and 3 analyses, we will examine the effect of sex (a biological variable) on study outcomes.**

Table 2. Quantitative Analyses: study variables and statistical methods

Research goals	Aim	Participants	Outcomes	Covariates/predictors	Statistical Methods
Evaluate Pulsewatch performance against gold standard	**2**	90 Pulsewatch users	pAF determined by 14-day cardiac event monitoring	Number and percentage of AF readings over 14 days	Logistic regression to produce area under the ROC curve (AUC)
Determine effects of patient characteristics on Pulsewatch performance	**2**	90 Pulsewatch users	Area under the curve	Patient characteristics (e.g., sex, hearing or visual impairment)	Chi-square test to compare AUC of 2 independent ROC curves
Examine Pulsewatch usability	**3**	30 Pulsewatch users	Indicator of watch daily use (yes vs. no) over 1 month	Sex, cognitive function, social support, depression, Barthel)	Mixed effects logistic regression model
Evaluate the impact of Pulsewatch use on key patientreported outcomes	**3**	30 Pulsewatch users and 30 controls	Stroke-related quality of life, anxiety & self-activation (Δ from baseline -1 mo)	Potential confounders (e.g. sex, medication adherence and disease-management selfefficacy)	t-test and linear regression model

Aim 2 Analysis: The analysis will focus on examining the performance of Pulsewatch worn for 9 hours daily for 14 days compared to 14-day cardiac monitoring, CardioKey by Biotel. A participant will be classified as having or not having pAF at the end of 14 days based on the gold standard. The total number of positive readings collected over 14 days from Pulsewatch (min 1 pulse every 5 minutes for 504 samples) will be used as the independent variable in a logistic regression to predict the binary outcome of pAF (yes vs. no). The area under the receiver operating characteristic (ROC) curve (AUC) will be calculated based on the results from the logistic regression to evaluate Pulsewatch performance for pAF screening. From the ROC curve, we can identify the cutoff point of total number of positive readings that produces the highest sensitivity and specificity combination. Similar analysis will be conducted using the % of positive readings among all readings collected from each participant as the independent variable. The 95% confidence intervals (CI) of the AUC will be calculated using formula given by Hanley and McNeil. We will then conduct exploratory analyses to examine whether characteristics affect watch performance over 14 days. For example, we will compare the area under 2 independent ROC curves of female vs. male participants (or vision impaired vs. not) using a chi-square test.

Aim 2 Power: Preliminary data suggest that Pulsewatch will have sensitivity, specificity and AUC of at least .9 to detect pAF. We calculate the width of 95% CI for AUC that ranges from .90 to .95 using the proposed sample size of 90 and an estimated rate of 20% patients with pAF. The width of the 95% CI ranges from .14 to .20. The proposed sample size will therefore give a quite precise estimate of Pulsewatch performance relative to the gold standard.

Aim 3 Analysis: The primary analysis focuses on app adherence use in 30 users asked to continue to use the watch and upload readings daily for 1 month. We will determine daily adherence (yes vs. no) based on whether or not daily smartwatch pulse recordings are present. We will examine whether participant characteristics, e.g., sex, cognitive impairment, or a high degree of stroke-related disability (Barthel Index), affect the likelihood of adherence over the 1 month study period. We will use a mixed effects logistic regression model including the participant as the random effect to capture the correlation among repeated measures from the same participant, uses participant characteristics as the fixed effects, and a binary indicator of daily adherence as the dependent variable. We will also examine the adherence time-trend by

including time (day) as a fixed effect in the model and patient as a random effect to estimate the slope of adherence over time. To examine whether the time trend varies by participant characteristics, we will include the interaction between characteristics and time in the model so that the slope of adherence over time can be estimated for each category of patient characteristic variables and be compared among the categories (e.g., pAF diagnosed vs. no pAF). Secondary analyses will examine rates of adherence to Pulsewatch ECG prompts and will compare Pulsewatch users to the usual care group with respect to stroke-related quality of life, anxiety, and self-activation. We will calculate the change score between baseline and 1 month as dependent variable. First, a *t*-test will be used to compare 2 study groups on the change score. A linear regression model, which will include group indicator (Pulsewatch users vs. usual care) as the independent variable and possible confounders (e.g., medication adherence and patient demographics) as covariates, to estimate adjusted group differences on the change scores.

Aim 3 Power: For the primary analysis, a power analysis was conducted comparing the time-averaged probability of adherence between 2 groups using a repeated measures regression model. We used 12 and 18 patients (total=30) in each of the groups (e.g., cognitively impaired vs. not), and type I error=5%. With 30 adherence measures for each patient and the probably of adherence in 1 group (e.g., cognitively impaired) = 50%, we will have 80% power to detect a difference in the adherence of 12%-16% in the other group when auto-correlation ranges from .2-.5). The difference in the probability of adherence decreases to 9%-12% when the probability in 1 group=80%. Therefore, our proposed n=30 for Pulsewatch users provides sufficient power to detect differences in the probability of adherence of 9%-16% between 2 groups (e.g., cognitively impaired vs. not). The secondary analysis power calculation was conducted comparing 30 users to 30 controls on continuous outcomes (e.g., anxiety) using a *t*-test. We will have a power of 80% to detect an effect size of .74 with 5% type I error.