Supplementary Online Content

Mark DB, Anstrom KJ, Sheng S, et al. Effect of catheter ablation vs medical therapy on quality of life among patients with atrial fibrillation: the CABANA randomized clinical trial [published online March 15, 2019]. *JAMA*. doi:10.1001/jama.2019.0692

Supplement 2. Statistical Analysis Plan

This supplementary material has been provided by the authors to give readers additional information about their work.

STATISTICAL ANALYSIS PLAN

Study Title:	The Catheter Ablation versus Antiarrhythmic Drug Therapy for Atrial Fibrillation: Quality of Life Study
Analysis Type:	Quality of Life
Analysis Plan Version Date:	5 May 2018
Analysis Plan Authors:	Daniel Mark, MD, MPH; Melanie Daniels, BA; J. David Knight, MS; Shubin Sheng, PhD; Kevin J. Anstrom, PhD

1. INTRODUCTION

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia problem in contemporary medicine. Available evidence suggests that AF has substantial adverse effects on quality of life (QOL).[1] Prior to designing the Catheter Ablation Versus Antiarrhythmic Drug Therapy for Atrial Fibrillation (CABANA) trial, we examined what was known about the impact of the AF ablation procedure being studied in CABANA on physical functioning and quality of life outcomes. A systematic review identified 49 studies published between 1988 and 2005 that examined the impact of AF on QOL.[2] The vast majority of these studies were small, involving less than 200 patients. Three of the largest randomized clinical trials (STAF, PIAF, RACE) found a greater improvement in QOL with rate control than with rhythm control. However, the AFFIRM trial (n=716), the largest at that time, found no difference between the two strategies. The SAFE-T trial recently reported that patients who achieved sinus rhythm regardless of therapeutic mechanism had significantly improved QOL and exercise capacity.[3] A randomized multicenter pilot study comparing AF ablation with medical therapy in 70 symptomatic AF patients found that ablation therapy improved multiple domains of QOL at 6 months relative to antiarrhythmic therapy, particularly involving general health, physical functioning, and social functioning.[4] A second randomized trial of amiodarone with or without AF ablation in 146 patients with chronic AF found that ablation reduced the severity score on the symptom checklist.[5] There were no large-scale trials or well done observational studies that defined the presence and magnitude of QOL benefits of catheter ablation of AF relative to an appropriate medical therapy control.

Since the start of CABANA, several additional studies have published results on the use of ablation therapy to treat atrial fibrillation. In the MANTRA-PAF trial, 294 patients with paroxysmal AF were randomly assigned to ablation or antiarrhythmic drug therapy. The cumulative burden of AF over two years did not differ between treatment groups in this trial, but the ablation group had fewer patients with AF and there was a trend toward greater improvement in the Short Form (36) Health Survey (SF-36) physical component summary score at 24 months.[6] QOL scores did not change from 2-year follow-up to 5-year follow-up, and there were no statistically significant between-group differences.[7] The RAAFT-2 trial found no difference in EQ-5D scores at 12 months between groups assigned to either ablation or medical therapy.[8] The ThermoCool AF trial randomized 167 patients with symptomatic AF unresponsive to at least one antiarrhythmic drug in a 2:1 ratio to catheter ablation or medical therapy.[9] The primary comparison between treatment arms showed very large improvements in the SF-36 scales through 9 months of study follow-up. Symptom frequency and severity decreased more than 50% in patients treated with ablation. The APAF trial also tested ablation therapy versus antiarrhythmic therapy in 198 patients with paroxysmal AF unresponsive to at least one antiarrhythmic drug.[10] At 4 years of follow-up, SF-36 scores were not different between groups by intention-to-treat, but most patients in the drug therapy arm had crossed over to ablation by that time.

CABANA is a multi-center randomized clinical trial designed to assess the safety and efficacy of percutaneous left atrial catheter ablation versus antiarrhythmic drug therapy in subjects who are at least 18 years of age and have new onset or under-treated paroxysmal, persistent, or long-standing persistent atrial fibrillation (AF) and warrant therapy for their arrhythmia. The study population consists of 2,204 subjects enrolled at 126 clinical sites over a period of approximately six years. Subjects were randomized in equal proportions (1:1) to receive either catheter ablation or drug therapy and followed at regular intervals for the duration of the study. The minimum length of

follow-up will be slightly less than 2 years, and the median duration of clinical follow-up will be approximately 4 years.

A prospective quality of life study is being conducted in the CABANA trial to compare QOL outcomes in subjects randomized to either medical therapy or catheter ablation. Operationally, the QOL study involves the collection of QOL data at randomization, and at Months 3, 6, and every 6 months thereafter until the end of study (or at the end-of-treatment visit for subjects withdrawing from the study prior to the end of study visit).

Eligibility criteria are described in the clinical study protocol. All subjects were eligible to participate in the QOL study. The data used for the QOL analyses will include data collected as part of the main study (demographic and baseline characteristics, and medical outcomes) and data collected specifically for the QOL study (subject-reported QOL assessments).

1.1. QOL Study Objectives

To compare health-related quality of life for catheter ablation as compared with drug therapy by intention-to-treat.

1.2. Study Design

Design Configuration	Two-arm randomized, parallel assignment, unblinded study
Treatment Groups	Group 1: left atrial ablation Group 2: rate or rhythm control drug therapy
Key Eligibility Criteria	All subjects enrolled in the study were included in the QOL study.

2. TYPE OF PLANNED ANALYSIS

This Statistical Analysis Plan (SAP) describes the statistical design, derivations, and statistical procedures for the final analysis of the QOL study.

Patients completed a battery of QOL questionnaires designed to assess health status, atrial fibrillation-related symptoms, general and atrial fibrillation-specific physical and social activities, emotional well-being, and demographic items. The list of instruments used is shown in Table 1.

Table 1. QOL Study Questionnaires

Instrument	Used to Assess	Items	Scale of Measurement Analysis Methods To Be Used

Atrial Fibrillation Effect and Quality of Life (AFEQT)	Atrial Fibrillation-Specific Quality of Life	20	Interval
Mayo AF Symptom Inventory (MAFSI)	Atrial Fibrillation Symptoms	10	Interval
Duke Activity Status Index (DASI)	Physical Function	12	Interval
Toronto Atrial Fibrillation Severity Score	Atrial Fibrillation Symptoms	19	Interval
SF-36 Global Health	Global Health Utility	1	Ordinal
SF-36 General Health	General Health	5	Interval
SF-36 MHI-5	Overall Mental Health	5	Interval
SF-36 Physical Functioning	Overall Physical Functioning	10	Interval
SF-36 Social Functioning	Social Functioning	2	Interval
SF-36 Bodily Pain	Pain Consequences	2	Interval
SF-36 Vitality	Energy/Fatigue	4	Interval
SF-36 Role Physical	Physical Impact on Daily Activities	4	Interval
SF-36 Role Emotional	Emotional Impact on Daily Activities	3	Interval
EuroQoL (EQ-5D-3L)	Health State	5	Interval
EQ-5D Visual Analog Scale (EQ-5D VAS)	Health Utility	1	Interval
Stanford Presenteeism Scale (SPS-6)	Work Impact	6	Interval
Work Productivity and Activity Impairment Questionnaire (WPAI)	Work Impact	5	Interval

Additional information was collected on work status, educational background, and marital status.

3. GENERAL CONSIDERATIONS FOR DATA ANALYSES

3.1. Analysis Population

The analysis population for the QOL study will be all randomized patients.

3.2. Data Sources

The enrolling clinical sites complete a series of electronic Case Report Forms (eCRF) for collection of the clinical data on each subject randomized, including demographic, clinical, and quality of life forms at enrollment and at specified follow-up intervals (the follow-up intervals are detailed in the study protocol), event forms for death, neurological events, bleeding, cardiac arrest, and other ancillary forms as required. Additional details on clinical data collection and data management are provided in the main CABANA SAP.

The quality of life data collection involves a baseline phase and a follow-up phase. At baseline, the coordinators at each enrolling site were responsible for collection of all the baseline QOL data forms using structured interviews: Full QOL questionnaire, EQ-5D worksheet, and MAFSI worksheet. Each coordinator underwent training for this task by the EQOL CC at DCRI before beginning this data collection. Baseline QOL Questionnaire data was entered into a secure Access 2010 database at the DCRI EQOL CC. The baseline EQ-5D and MAFSI worksheets were entered by site staff into the InForm electronic data capture (EDC) system.

In the original design of CABANA, all enrollment was envisioned to come from North America and the plan for follow-up QOL data collection was to have the Call Center at the DCRI EQOL CC perform the scheduled structured interviews by phone. This plan had to be modified when enrollment was extended internationally so that for sites outside North America, follow-up interviews were conducted by the site coordinators of the individual international sites and then transmitted via secured facsimilie or email to the EQOL CC. Follow-up QOL questionnaires were entered into the EQOL CC Access database. All sites continued to enter EQ-5D and MAFSI worksheets into InForm during follow-up intervals.

All QOL data collected in the study are being managed and analyzed at the CABANA EQOL CC in the DCRI. The QOL questionnaires are carefully reviewed for completeness and run through data quality checks. The InForm eCRF data are imported into the EQOL CC Access database and then downloaded as raw SAS data files, and further review and checking of the data occur. The raw SAS data, analysis datasets, and analysis programs are stored on the secure DCRI statistical server. Final analyses will be performed at the DCRI using SAS version 9.4 or higher (SAS Institute Inc., Cary, NC).

3.3. Treatment Groups

Subjects will be grouped for analyses according to the randomized treatment assignment using the principle of intention to treat.

3.4. Data Analysis

The co-primary endpoints for the QOL study are the AFEQT and MAFSI.

Statistical comparisons will be performed using linear models that are appropriate for the outcome variables. There will be no adjustments for multiple comparisons. All secondary QOL endpoint comparisons (Table 2) are considered to serve the role of supporting and enhancing our understanding of the patient's perspective on the two treatments tested in CABANA.

Table 2. Secondary Endpoints

DASI
Toronto Atrial Fibrillation Severity Scale
SF-36 Global Health Utility
SF-36 General Health
SF-36 Mental Health Inventory-5 (MHI-5)
SF-36 Physical Functioning
SF-36 Social Functioning
SF-36 Bodily Pain
SF-36 Vitality
SF-36 Role Physical
SF-36 Role Emotional
EQ-5D-3L
EQ-VAS
Stanford Presenteeism Scale (SPS-6)
Work Productivity and Activity Impairment (WPAI)

3.5. Missing Data

Extensive procedural efforts were made during CABANA to minimize the amount of missing data on the QOL questionnaires. Study personnel verified that any missing responses in the questionnaires were intentional on the subject's part. Subjects who discontinued drug therapy were encouraged to continue in all study assessments as scheduled and will be included in the analyses.

In preliminary analysis work, we will evaluate patterns of missingness as follows:

- compare subjects with missing data versus subjects without missing data by study center, demographics, and other relevant subject characteristics; we will also examine narrative reasons for missing data when available to see if the missingness can be classified as administrative or disease-related
- evaluate the reasons and time to premature study discontinuation by treatment group among the QoL analysis population; and
- examine the frequency of subjects, by treatment group, in relation to the observed permutations of missingness across time points.

Rules for handling missing items within questionnaire domains (ie, the proportion of items that can be missing before an endpoint is treated as missing) and the statistical approaches used to address instances where the entire domain score or instrument for a subject is missing are provided in Section 6.

3.6. Follow-up Contact Windows

Full QOL questionnaires were collected at randomization and months 3, 12, and every 12 months thereafter (or at the end of treatment visit for subjects withdrawing from the study prior to the end of study visit). Brief QOL questionnaires were collected at month 6 and every 12 months thereafter. Abbreviated proxy QOL questionnaires were collected for incapacitated patients.

The completion of an end-of-treatment contact could result in a subject having more than 1 set of questionnaires categorized to a given study contact. In those instances, the full questionnaire completed closest to the target date of the expected study contact will be used in the analyses, using the earlier questionnaire in case of a tie.

3.7. Timing of Analysis

The analysis of the unblinded data for the QOL study will be conducted following the lock of the clinical database. The need for more than one database lock is not anticipated for this study. Therefore, data handling and blinding issues relevant to studies with multiple, planned database locks are not applicable.

3.8. "On Treatment Analysis"

The intention-to-treat analyses in this trial will constitute the primary analyses and will serve as the standard for interpreting treatment differences in the key clinical outcomes. However, because a number of patients in the drug arm may cross over to receive ablation during the trial, and some patients randomized to ablation may not undergo the procedure, we will supplement the intent-to-treat comparisons with "on-treatment" comparisons. The "on-treatment" analysis will involve a comparison of patients who received ablation (even if originally assigned to the drug arm) versus those who did not.

3.9. Per-Protocol Analysis

An analysis will also be performed to compare the primary endpoint and the key secondary clinical outcomes of the two treatment strategies among the subset of patients who fully satisfied the inclusion/exclusion criteria and received the treatment to which they were randomly allocated. This analysis will include patients randomized to the drug arm who were treated with drug therapy, and patients randomized to the ablation arm who underwent the ablation within 6 months following enrollment in the trial. The follow-up of drug-arm patients who crossed over to ablation will be censored at the time of the ablation. Patients randomized to the ablation arm who were not ablated within 6 months will be excluded from this analysis.

4. BASELINE QOL DATA

Baseline responses/scores on the QOL endpoints will be summarized by treatment group and overall. The endpoints will be summarized using the standard 5 number summary – mean, standard deviation, median, 25^{th} and 75^{th} percentiles.

5. SUBJECT CHARACTERISTICS AND SUBJECT DISPOSITION DATA

As discussed in section 3.5, subjects with missing data will be compared with subjects without missing data by study center, demographics, and other key subject characteristics. A Pearson chi-square test for categorical variables and a Wilcoxon rank sum test for continuous variables will be used to help gauge any differences between the two cohorts.

Subject disposition among the QOL analysis population will be summarized, including reasons for premature study discontinuation. In addition, the frequency of subjects in relation to the observed permutations of missingness across the expected time points will be summarized by treatment group.

6. ENDPOINT ANALYSES

6.1. Primary QOL Endpoints

The two coprimary endpoints for the QOL comparison in CABANA are the Atrial Fibrillation Effect and Quality of Life Scale (AFEQT) and the Mayo AF-Specific Symptom Inventory (MAFSI). The primary analysis of these endpoints will focus on the estimated treatment effect at 12 months. Comparisons at other time points and overall (integrating across all follow-up) will be considered secondary comparisons.

6.1.1. Primary QOL Endpoint Definitions

Atrial Fibrillation Effect and Quality of Life Scale (AFEQT)

The **AFEQT**[11] is a validated 20-item atrial fibrillation-specific, health-related quality of life questionnaire designed to assess the impact of atrial fibrillation on patients HRQOL and possibly changes with treatment. The AFEQT has an Overall Score and subscale scores in three domains: symptoms, daily activities, and treatment concern. Two questions regarding satisfaction with health care providers and with treatment are not included in the Overall AFEQT Score and were not collected for the CABANA study.

General Scoring Information
The responses on the AFEQT are scored on a 1 to 7 Likert scale, where 1 = ”Not at all…” to 7 = ”Extremely…”.

If questionnaire says ‘no AF symptoms’ use response options :
o “Not Bothered at All OR I Did Not Have This Symptom”
o “Not At All Limited”
o “Not At All Bothered”
o “No Difficulty At All”
o “Not Applicable”
o “Strongly Disagree” (work questions)

FOR intervals > 18 months: If patient states " I haven't had Atrial Fibrillation >1 year ago OR say "I was never aware of having atrial fibrillation", use skip pattern:
Brief Follow-Up QX (18, 30 and 42 Mo intervals) skip qxs # 3-7
Full Follow-up QX (24, 36, and 48 Mo intervals) skip qxx # 15-19
Proxy QX (18, 24, 30, 36, 42, 48 Mo intervals) skip qxs # 15-17

Overall AFEQT HRQOL Score
The AFEQT HRQOL Score is calculated based on the following formula:

$$100 - \frac{(\text{sum of severity for all questions answered} - \text{number of questions answered}) \times 100}{(\text{total number of questions answered} \times 6)}$$

Subscale Scores
Subscale scores are computed similarly to the Overall AFEQT Score from each subscale used to generate its own score.

The 18 questions are grouped into 3 functional subscales as described below:

Table 5: AFEQT Subscales

Subscales	**Questions on the Full QOL Questionnaire**
Symptoms	15a, 15b, 15c, and 15d
Daily Activities	21a, 21b, 24a, 24b, 24c, 24d, 24e, and 24f
Treatment Concern	22d, 22e, 23a, 23b, 23c, and 23d

Interpretation
Overall or subscale scores range from 0 to 100. A score of 0 corresponds to complete disability (or responding "extremely" limited, difficult, or bothersome to all questions answered), while a score of 100 corresponds to no disability (or responding "not at all" limited, difficult, or bothersome to all questions answered).

For example, if a patient answered all "1" for the Symptoms subscale, the subscale score would be 100 – [(4 – 4) / 4 x 6] x 100 = 100 – [0 / 24] x 100 = 100 or patient has no disability.

Conversely, if a patient answered all "7" for the Symptoms subscale, the subscale score would be 100 – [(28 – 4) / 4 x 6] x 100 = 100 – [24 / 24] x 100 = 0 or patient is extremely limited.

Handle Missing Items:
At least 50% of completed responses for each domain are required to calculate a meaningful score.

Mayo AF-Specific Symptom Inventory (MAFSI)

We used a modified **MAFSI**[12] questionnaire comprised of a 10-item atrial fibrillation symptom checklist asking for both frequency and severity of each symptom. The frequency of symptoms

over the past month is recorded as 0 (never), 1 (rarely), 2 (sometimes), 3 (often), and 4 (always). Items are then summed for a total frequency score. The severity of each symptom is recorded as 1 (mild), 2 (moderate), and 3 (extreme). Items are then summed for the total severity score.

Handle Missing Items:
If four or more items are missing, the MAFSI endpoint will be considered missing.

Interpretation
Total frequency score ranges from 0 to 40 with a score of 0 indicating no atrial fibrillation symptoms. Total severity score ranges from 0 to 30, with a score of 30 indicating the most severe symptoms.

6.1.2. Analysis of Primary QOL Endpoints

The primary analysis for the AFEQT and MAFSI assessments will be performed using the intention-to-treat principle on the QOL data analysis set. The AFEQT and MAFSI endpoints will be analyzed using a repeated-measures mixed model with the baseline score as a covariate, Month 3, Month 12, Month 24, Month 36, Month 48, and Month 60 responses included as outcome variables, and time as a fixed variable.

Restricted maximum likelihood estimation will be used to model all available data from each subject without imputing missing values. An unstructured covariance matrix will be used.

Point estimates for each treatment group and treatment group mean differences (ablation – medical treatment) with 95% confidence intervals (CIs) will be generated for each time point. The primary assessment will be based on the treatment group difference at Month 12. Additional analyses will examine the treatment effect at Months 3, 24, 36, 48, and 60. Additionally, the treatment effect will be averaged across the 6 follow-up time points. The estimated treatment difference and 95% CIs will be obtained using the ESTIMATE Statement in SAS PROC MIXED.

6.2. Secondary QOL Endpoints

6.2.1. Secondary QOL Endpoint Definitions

Duke Activity Status Index (DASI)

The **DASI**[13] is a validated 12-item questionnaire that assesses general physical functioning; activities range in physical demands from self-care to strenuous sports. The DASI is comprised of questions 2 through 13 on the QOL CRF.

The possible responses to each item are 1 = Yes, with no difficulty, 2 = Yes, but with some difficulty OR I couldn't do this, 3 = Don't do this for other reasons. Each item answered 'Yes, with no difficulty' will be assigned a weighted score corresponding to the average amount of metabolic output implied in its performance; see Table 6 below. Items answered "Yes, but with some difficulty OR I couldn't do this" or "Don't do this for other reasons" will be assigned a score of 0. If more than 4 items are missing, the DASI endpoint will be considered missing.

Otherwise, weighted scores will be summed to achieve a total score. The total score can range from 0 to 58.2; lower scores indicate worse physical functioning. The total score will be used in the analyses. For an individual patient, a clinically significant change is considered to be 4 points or more.

Table 6. Weighted Scores Assigned to DASI items

DASI Item #	eCRF Question #	Activity	Yes, with no difficulty	Yes, but with some difficulty OR I couldn't do this	Don't do this for other reasons
1	2	Could you take care of yourself (eating, dressing, bathing or using the toilet)?	2.75	0	0
2	3	Could you walk indoors, such as around your house?	1.75	0	0
3	4	Could you walk a block or two on level ground?	2.75	0	0
4	5	Could you climb a flight of stairs or walk up a hill?	5.50	0	0
5	6	Could you run a short distance?	8.00	0	0
6	7	Could you do light work around the house like dusting or washing dishes?	2.70	0	0
7	8	Could you do moderate work around the house like vacuuming, sweeping floors or carrying in groceries?	3.50	0	0
8	9	Could you do heavy work around the house like scrubbing floors or lifting and moving heavy furniture?	8.00	0	0
9	10	Could you do yard work like raking leaves, weeding or pushing a power mower?	4.50	0	0
10	11	Could you have sexual relations?	5.25	0	0
11	12	Could you participate in moderate recreational activities like golf, bowling, dancing, doubles tennis or throwing a baseball or football?	6.00	0	0
12	13	Could you participate in strenuous sports like swimming, singles tennis, football, basketball or skiing?	7.50	0	0

The University of Toronto Atrial Fibrillation Severity Scale (AFSS)

The Atrial Fibrillation Severity Scale is a 19-item disease-specific questionnaire used to assess AF-related symptoms, health care utilization, and the frequency, duration, and severity of AF

episodes.[12] Questions on the CABANA QOL CRF corresponding to the below scoring instructions are as follows:

Toronto AFSS Question #	CABANA QOL Full Questionnaire #
5	16
6	17
7	19
8	18

The **Total AF Burden** score is obtained by combining measures of frequency (question #5), duration (question #6), and patient perceived severity (the average of questions #7 & #8).

Question #5 has responses ranging from 1-11, and patients that respond "less than once a year" are given a score of 10 instead of 11 for the purpose of this calculation. For question #6, the score for each patient is divided by 8 and multiplied by 10 so that each patient will have a value for this question that falls in the range of 1-10.

Once that is completed, both questions #5 and #6 are reverse coded (ie. for question #5, "continuous" should be reverse coded so that it =10, "more than twice a day"=9, etc).

Total AF Burden = AF frequency + AF duration + AF severity. Each of the 3 measures contribute equally to the AF burden score, and each measure ranges from 1-10 to yield Total AF Burden scores ranging from 3-30.

Interpretation
Higher scores indicate greater AF burden.

The **AF Severity** score is calculated as the arithmetic mean of questions #7 & #8. Score range from 1 to 10.

Interpretation
Higher scores indicate greater severity.

Handle Missing Items:
If one or more items are missing, the Total AF Burden Score AFSS endpoint will be considered missing.

The Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36)

The SF-36[14] is a 36-item questionnaire used to provide a detailed assessment of functioning and well-being from a generic health status perspective. The instrument provides an Global Health Utility and 8 subscales. The SF-36 asks patients to recall health status “during the past 4 weeks.” Scores range from 0 to 100 with higher scores representing better health status. A clinically significant change for a patient using this scoring system has not been established but can be approximated by a ¼ standard deviation, or 5 points or more.

The **SF-36 Global Health Utility** is a 1-item question from the SF-36 General Health scale that assesses general health perception. The GHU is question 1 on the Full QOL Questionnaire. The possible ordinal responses to the question are 1 = Excellent, 2 = Very Good, 3 = Good, 4 = Fair, and 5 = Poor. If the response is missing, then the endpoint will be considered missing.
The GHU will be analyzed as an ordinal outcome measure. A nonparametric approach will be used to compare differences between treatment groups at each time point. Missing data will be imputed as the worst possible value in the non-parametric analysis.

The **SF-36 General Health Scale** is a 5-item questionnaire that assesses overall health status. The general health items are questions 1, 38a, 38b, 38c, and 38d on the Full QOL Questionnaire. The possible responses to question 1 are 1 = Excellent, 2 = Very Good, 3 = Good, 4 = Fair, and 5 = Poor. The possible responses to questions 38a—38d are 1 = Definitely true, 2 = Mostly true, 3 = Don't know, 4 = Mostly false, and 5 = Definitely false. The transformed general health scale score will be used for the analyses. Transformed scores range from 0 to 100; lower scores indicate worse general health status.

The procedures below will be followed to obtain a transformed general health score:

- Assign a final value for each item:
  - The pre-coded values for question 1, 38b, and 38d will be reverse coded to obtain the final value (e.g., a response of '1 = Excellent' will be assigned a value of 5; '2 = Very Good' will be assigned a value of 4).
  - The pre-coded values for items 38a and 38c will be retained as the final item value (e.g., a response of '1 = Definitely true' will be assigned a value of 1; '2 = Mostly true' will be assigned a value of 2).
  - This process will result in higher values for each item indicating better general health.
- Handle missing items:
  - The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 3 of the 5 items, then a person-specific estimate for any missing items will be imputed. If more than 2 items are missing, then the general health endpoint will be considered missing.
- Compute the raw scale score:
  - After recoding and imputing missing values, the final item values for all 5 items will be summed to achieve a raw scale score.
- Transform the raw scale score:
  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.

- For the general health scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 5) ÷ 20] * 100

- The transformed score represents the percentage of the total possible score achieved.

The **SF-36 Mental Health Inventory – 5 (MHI-5) Scale** is a 5-item questionnaire that assesses mental health including depression and anxiety. The mental health items are questions 36b, 36c, 36d, 36f, and 36h on the Full QOL Questionnaire. The possible responses to the 5 items are 1 = All of the time, 2 = Most of the time, 3 = Some of the time, 4 = A little of the time, and 5 = None of the time. The transformed mental health scale score will be used for the analyses. Transformed scores range from 0 to 100; lower scores indicate worse mental health status.

The procedures below will be followed to obtain a transformed mental health score:

- Assign a final value for each item:

  - The pre-coded values for items 36d and 36h will be reverse coded to obtain the final value (e.g., a response of '1 = All of the time' will be assigned a value of 5; '2 = Most of the time' will be assigned a value of 4).

  - The pre-coded values for items 36b, 36c, and 36f will be retained as the final item value (e.g., a response of '1 = All of the time' will be assigned a value of 1; '2 = Most of the time' will be assigned a value of 2).

  - This process will result in higher values for each item indicating better mental health.

- Handle missing items:

  - The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 3 of the 5 items, then a person-specific estimate for any missing items will be imputed. If more than 2 items are missing, then the mental health endpoint will be considered missing.

- Compute the raw scale score:

  - After recoding and imputing missing values, the final item values for all 5 items will be summed to achieve a raw scale score.

- Transform the raw scale score:
  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.

  - For the mental health scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 5) ÷ 20] * 100

  - The transformed score represents the percentage of the total possible score achieved.

The **SF-36 Physical Functioning Scale** is a 10-item questionnaire that assesses physical functioning. The physical functioning items are questions 39a, 39b, 39c, 39d, 39e, 39f, 39g, 39h, 39i, and 39j on the Full QOL Questionnaire. The possible responses to the 10 items are 1 = Yes, limited a lot, 2 = Yes, limited a little, and 3 = No, not limited at all. The transformed physical functioning scale score will be used for the analyses. Transformed scores range from 0 to 100; lower scores indicate worse physical functioning status.

The procedures below will be followed to obtain a transformed physical functioning score:

- Assign a final value for each item:
  - The pre-coded values for all 10 items will be retained as the final item value (e.g., a response of '1 = Yes, limited a lot' will be assigned a value of 1; '2 = Yes, limited a little' will be assigned a value of 2).
  - This process will result in higher values for each item indicating better physical function.
- Handle missing items:
  - The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 5 of the 10 items, then a person-specific estimate for any missing items will be imputed. If more than 5 items are missing, then the physical functioning endpoint will be considered missing.
- Compute the raw scale score:
  - After recoding and imputing missing values, the final item values for all 10 items will be summed to achieve a raw scale score.
- Transform the raw scale score:
  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.
  - For the physical functioning scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 10) ÷ 20] * 100
  - The transformed score represents the percentage of the total possible score achieved.

The **SF-36 Social Functioning Scale** is a 2-item questionnaire that assesses the effect of physical health or emotional problems on social activities. The social functioning items are questions 33 and 37 on the Full QOL Questionnaire. The possible responses to the extent of limitation are 1 = Not at all, 2 = Slightly, 3 = Moderately, 4 = Quite a bit, and 5 = Extremely. The possible responses to the duration of limitation are 1 = All of the time, 2 = Most of the time,

3 = Some of the time, 4 = A little of the time, and 5 = None of the time. The transformed social functioning scale score will be used for the analyses. Transformed scores range from 0 to 100; lower scores indicate less social functioning.

The procedures below will be followed to obtain the transformed social functioning score:

- Assign a final value for each item:
  - The pre-coded values for item 33 will be reverse coded to obtain the final item value (e.g., a response of '1 = Not at all' will be assigned a value of 5; '2 = Slight' will be assigned a value of 4).
  - The pre-coded values for items 37 will be retained as the final item value (e.g., a response of '1 = All of the time' will be assigned a value of 1; '2 = Most of the time' will be assigned a value of 2).
  - This process will result in higher values for each item indicating better social functioning.
- Handle missing items:
  - The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 1 of the 2 items, then a person-specific estimate for any missing items will be imputed. If both items are missing, then the social functioning endpoint will be considered missing.
- Compute the raw scale score:
  - After recoding and imputing missing values, the final item values for both items will be summed to achieve a raw scale score.
- Transform the raw scale score:
  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.
  - For the social functioning scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 2) ÷ 8] * 100
  - The transformed score represents the percentage of the total possible score achieved.

The **SF-36 Bodily Pain Scale** is a 2-item questionnaire that assesses the magnitude and effect of bodily pain over the past 4 weeks. The social functioning items are questions 34 and 35 on the Full QOL Questionnaire. The possible responses to the severity of pain are 1 = None, 2 = Very mild, 3 = Mild, 4 = Moderate, 5 = Severe, and 6 = Very severe. The possible responses to the effect of pain on work activity are 1 = Not at all, 2 = A little bit, 3 = Moderately, 4 = Quite a bit,

and 5 = Extremely. The transformed social functioning scale score will be used for the analyses. Transformed scores range from 0 to 100; lower scores indicate worse pain status.

The procedures below will be followed to obtain the transformed bodily pain score:

- Assign a final value for each item:
  - The pre-coded values for both items 34 and 35 will be reverse coded to obtain the final item value (e.g., a response of '1 = None' will be assigned a value of 6 for item 34; '2 = Very mild' will be assigned a value of 5 for item 34).
  - This process will result in higher values for each item indicating better pain status.
- Handle missing items:
  - The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 1 of the 2 items, then a person-specific estimate for any missing items will be imputed. If both items are missing, then the bodily pain endpoint will be considered missing.
- Compute the raw scale score:
  - After recoding and imputing missing values, the final item values for both items will be summed to achieve a raw scale score.
- Transform the raw scale score:
  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.
  - For the bodily pain scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 2) ÷ 9] * 100
  - The transformed score represents the percentage of the total possible score achieved.

The **SF-36 Vitality Scale** is a 4-item questionnaire that assesses vitality including energy level and fatigue. The vitality items are questions 36a, 36e, 36g, and 36i on the Full QOL Questionnaire. The possible responses to the 4 items are 1 = All of the time, 2 = Most of the time, 3 = Some of the time, 4 = A little of the time, and 5 = None of the time. The transformed vitality scale score will be used for the analyses. Transformed scores range from 0 to 100; lower scores indicate less vitality.

The procedures below will be followed to obtain the transformed vitality score:

- Assign a final value for each item:

- The pre-coded values for items 36a and 36e will be reverse coded to obtain the final item value (e.g., a response of '1 = All of the time' will be assigned a value of 5; '2 = Most of the time' will be assigned a value of 4).

- The pre-coded values for items 36g, and 36i will be retained as the final item value (e.g., a response of '1 = All of the time' will be assigned a value of 1; '2 = Most of the time' will be assigned a value of 2).

- This process will result in higher values for each item indicating more vitality.

- Handle missing items:

  - The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 2 of the 4 items, then a person-specific estimate for any missing items will be imputed. If more than 2 items are missing, then the vitality endpoint will be considered missing.

- Compute the raw scale score:

  - After recoding and imputing missing values, the final item values for all 4 items will be summed to achieve a raw scale score.

- Transform the raw scale score:

  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.

  - For the vitality scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 4) ÷ 16] * 100

  - The transformed score represents the percentage of the total possible score achieved.

The **SF-36 Role-Physical Scale** is a 4-item questionnaire that assesses physical limitations on daily activities. The role-physical items are questions 31a, 31b, 31c, and 31d on the Full QOL Questionnaire. The possible responses to the 4 items are 1 = All of the time, 2 = Most of the time, 3 = Some of the time, 4 = A little of the time, and 5 = None of the time. The transformed role-physical scale score will be used for the analyses. Transformed scores range from 0 to 100; higher scores indicate fewer physical limitations on daily activities.

The procedures below will be followed to obtain the transformed role-physical score:

- Assign a final value for each item:

  - The pre-coded values for all 4 items will be retained as the final item value (e.g., a response of '1 = All of the time' will be assigned a value of 1; '2 = Most of the time' will be assigned a value of 2).

- This process will result in higher values for each item indicating fewer physical limitations on daily activities.

- Handle missing items:

  - The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 2 of the 4 items, then a person-specific estimate for any missing items will be imputed. If more than 2 items are missing, then the role-physical endpoint will be considered missing.

- Compute the raw scale score:

  - After recoding and imputing missing values, the final item values for all 4 items will be summed to achieve a raw scale score.

- Transform the raw scale score:

  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.

  - For the role-physical scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 4) ÷ 16] * 100

  - The transformed score represents the percentage of the total possible score achieved.

The **SF-36 Role-Emotional Scale** is a 3-item questionnaire that assesses emotional limitations on daily activities. The role-emotional items are questions 32a, 32b, and 32c on the Full QOL Questionnaire. The possible responses to these items are 1 = All of the time, 2 = Most of the time, 3 = Some of the time, 4 = A little of the time, and 5 = None of the time. The transformed role-emotional scale score will be used for the analyses. Transformed scores range from 0 to 100; higher values indicate fewer emotional problems limiting daily activities.

The procedures below will be followed to obtain the transformed bodily pain score:

- Assign a final value for each item:

  - The pre-coded values for all 3 items will be retained as the final item value (e.g., a response of '1 = All of the time' will be assigned a value of 1; '2 = Most of the time' will be assigned a value of 2).

  - This process will result in higher values for each item indicating fewer emotional limitations on daily activities.

- Handle missing items:

- The "half-scale" rule for imputing missing scores will be applied. That is, if a subject answered at least 2 of the 3 items, then a person-specific estimate for any missing items will be imputed. If all 3 items are missing, then the role-emotional endpoint will be considered missing.

- Compute the raw scale score:

  - After recoding and imputing missing values, the final item values for both items will be summed to achieve a raw scale score.

- Transform the raw scale score:

  - The raw score is transformed by subtracting the lowest possible raw score from the actual raw score, dividing by the possible raw score range, and multiplying by 100. This converts the lowest and highest possible scores to 0 and 100, respectively.

  - For the role-emotional scale, the following derivation will be used: Transformed Scale Score = [(actual raw score – 3) ÷ 12] * 100

  - The transformed score represents the percentage of the total possible score achieved.

EuroQoL

The **EQ-5D-3L**[15] is a 5-item questionnaire that measures health status on 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. This instrument was collected with the clinical CRF.

There are 3 levels of response for each dimension: 1) no problems, 2) some problems, 3) extreme problems. Level one is coded as 1; level two is coded as 2; and level three is coded as 3. For example, the possible responses for the mobility dimension are 1 = I have no problems in walking about, 2 = I have some problems in walking about, 3 = I am confined to bed. A unique health state for each subject is achieved by combining the level from each of the 5 dimensions, which is referred to as a 5-digit code. A total of 243 possible health states are defined in this way. If any of the 5 items are missing, then the EQ-5D-3L endpoint will be considered missing. Otherwise, the descriptive health states will be converted into a summary index score by applying an algorithm that assigns weights to each level in each dimension.

The **EQ VAS** is a vertical, visual analogue scale where the ends are labeled 'best imaginable health state' and 'worst imaginable health state.' This instrument was collected with the clinical CRF.

If the VAS score is missing, the EQ VAS endpoint will be considered missing. The VAS score ranges from 0 to 100; lower scores indicate worse health.

Stanford Presenteeism Scale (SPS-6)

The **SPS-6**[16] is a validated 6-item questionnaire used to measure the impact of a worker's perceived ability to concentrate on work tasks despite the distractions of health impairments and pain. It is captured by questions 25a to 25f on the Full QOL Questionnaire. The recall period is one month.

Responses are graded on a Likert 5-item response scale ranging from "Disagree strongly" to "Agree strongly."

Items 25a, 25c, and 25d are scored as follows: "Disagree strongly" = 5; "Disagree somewhat" = 4; "Uncertain" = 3; "Agree somewhat" = 2; and "Agree strongly" = 1. Items 25b, 25e, and 25f are scored as follows: "Disagree strongly" = 1; "Disagree somewhat" = 2; "Uncertain" = 3; "Agree somewhat" = 4; and "Agree strongly" = 5.

Then scores are summed for the SPS-6 total score. Scores can range from 6 to 30, with lower scores indicating lower presenteeism, and higher scores indicating higher presenteeism.

Handle Missing Items:
If one or more items are missing, the SPS-6 endpoint will be considered missing.

Work Productivity and Activity Impairment Scale (WPAI)

The **WPAI**[17] is a 6-item questionnaire that assesses the amount of absenteeism, presenteeism, and daily activity impairment attributable to health. It is comprised of questions 26 through 30 on the Full QOL Questionnaire. Outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity (i.e., worse outcomes) as follows:

WPAI:GH
WPAI General Health outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, i.e., worse outcomes, as follows:

WPAI Questionnaire	CABANA Full QOL Questionnaire Number
1 = currently employed	26
2 = hours missed due to health problems	27
3 = hours missed other reasons	28
4 = hours actually worked	29
5 = degree health affected productivity while working	30
6 = degree health affected regular activities	N/A

Scores:
Multiply scores by 100 to express in percentages.

Percent work time missed due to health: Q2/(Q2+Q4)

Percent impairment while working due to health: Q5/10

Percent overall work impairment due to health:
Q2/(Q2+Q4)+[(1-(Q2/(Q2+Q4)))x(Q5/10)]

Handle Missing Items:
If one or more items are missing, the WPAI:GH endpoint will be considered missing.

6.2.2. Secondary QOL Endpoint Analysis Methods

The DASI, AFSS, SF-36, EQ-5D, SPS-6, and WPAI endpoints will be analyzed as interval-scale outcome measures using the previously described repeated measures mixed model. The treatment effects will be examined at all assessment time points.

6.3. Subgroup Analyses

- Age (<65 , 65 to 74, and ≥75 years)
- Sex (male vs. female)
- Race (white vs. racial minorities)
- AF type (paroxysmal vs. persistent, or long-standing persistent)
- Years since onset of AF (>1 vs ≤1)
- Days from most recent qualifying episode of atrial fibrillation to enrollment (>12 vs ≤12 days)
- NYHA Heart Failure Class at enrollment (no heart failure or Class I vs. ≥ Class II)
- History of congestive heart failure (yes vs. no)
- Structural heart disease (present vs. absent)
- Hypertension (present vs. absent)
- Hypertension with LVH (present vs. absent)
- CHADS2 (0 or 1 vs >1)
- CHA_2DS_2-VASc score (0 or 1 vs >1)
- Sleep Apnea (present vs. absent)
- Family history of atrial fibrillation (yes vs. no)
- Obesity (BMI ≥30 vs. < 30)
- Left ventricular ejection fraction (LVEF) (≤35 vs >35)
- North American vs. other international sites

The main subgroups examined in the QOL data analysis will be those identified as of highest interest for the clinical analysis (see the CABANA clinical SAP). The estimated treatment effect of catheter ablation within each of the subgroup sets listed above will be examined. The interaction between subgroup and treatment will be evaluated. The subgroup analyses will be conducted on the primary model/analysis for each of the two co-primary QOL endpoints (AFEQT and MAFSI). The subgroup analyses will not be repeated in the various sensitivity analyses or for the exploratory endpoints unless the co-primary or clinical analyses identify important subgroup effects.

7. REFERENCES

1. Dorian P, Jung W, Newman D, Paquette M, Wood K, Ayers GM, Camm J, Akhtar M, Luderitz B. The impairment of health-related quality of life in patients with intermittent atrial fibrillation: implications for the assessment of investigational therapy. J Am Coll Cardiol. 2000;36(4):1303-9. Epub 2000/10/12. PubMed PMID: 11028487.
2. Thrall G, Lane D, Carroll D, Lip GY. Quality of life in patients with atrial fibrillation: a systematic review. Am J Med. 2006;119(5):448 e1-19. Epub 2006/05/03. doi: 10.1016/j.amjmed.2005.10.057. PubMed PMID: 16651058.
3. Singh SN, Tang XC, Singh BN, Dorian P, Reda DJ, Harris CL, Fletcher RD, Sharma SC, Atwood JE, Jacobson AK, Lewis HD, Jr., Lopez B, Raisch DW, Ezekowitz MD, Investigators S-T. Quality of life and exercise performance in patients in sinus rhythm versus persistent atrial fibrillation: a Veterans Affairs Cooperative Studies Program Substudy. J Am Coll Cardiol. 2006;48(4):721-30. Epub 2006/08/15. doi: 10.1016/j.jacc.2006.03.051. PubMed PMID: 16904540.
4. Wazni OM, Marrouche NF, Martin DO, Verma A, Bhargava M, Saliba W, Bash D, Schweikert R, Brachmann J, Gunther J, Gutleben K, Pisano E, Potenza D, Fanelli R, Raviele A, Themistoclakis S, Rossillo A, Bonso A, Natale A. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of symptomatic atrial fibrillation: a randomized trial. JAMA. 2005;293(21):2634-40. Epub 2005/06/02. doi: 10.1001/jama.293.21.2634. PubMed PMID: 15928285.
5. Oral H, Pappone C, Chugh A, Good E, Bogun F, Pelosi F, Jr., Bates ER, Lehmann MH, Vicedomini G, Augello G, Agricola E, Sala S, Santinelli V, Morady F. Circumferential pulmonary-vein ablation for chronic atrial fibrillation. N Engl J Med. 2006;354(9):934-41. Epub 2006/03/03. doi: 10.1056/NEJMoa050955. PubMed PMID: 16510747.
6. Cosedis Nielsen J, Johannessen A, Raatikainen P, Hindricks G, Walfridsson H, Kongstad O, Pehrson S, Englund A, Hartikainen J, Mortensen LS, Hansen PS. Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation. N Engl J Med. 2012;367(17):1587-95. Epub 2012/10/26. doi: 10.1056/NEJMoa1113566. PubMed PMID: 23094720.
7. Nielsen JC, Johannessen A, Raatikainen P, Hindricks G, Walfridsson H, Pehrson SM, Englund A, Hartikainen J, Mortensen LS, Hansen PS, Investigators M-P. Long-term efficacy of catheter ablation as first-line therapy for paroxysmal atrial fibrillation: 5-year outcome in a randomised clinical trial. Heart. 2017;103(5):368-76. Epub 2016/08/28. doi: 10.1136/heartjnl-2016-309781. PubMed PMID: 27566295.
8. Morillo CA, Verma A, Connolly SJ, Kuck KH, Nair GM, Champagne J, Sterns LD, Beresh H, Healey JS, Natale A, Investigators R-. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of paroxysmal atrial fibrillation (RAAFT-2): a randomized trial. JAMA. 2014;311(7):692-700. Epub 2014/02/20. doi: 10.1001/jama.2014.467. PubMed PMID: 24549549.
9. Reynolds MR, Walczak J, White SA, Cohen DJ, Wilber DJ. Improvements in symptoms and quality of life in patients with paroxysmal atrial fibrillation treated with radiofrequency catheter ablation versus antiarrhythmic drugs. Circ Cardiovasc Qual Outcomes. 2010;3(6):615-23. Epub 2010/10/14. doi: 10.1161/CIRCOUTCOMES.110.957563. PubMed PMID: 20940250.
10. Pappone C, Vicedomini G, Augello G, Manguso F, Saviano M, Baldi M, Petretta A, Giannelli L, Calovic Z, Guluta V, Tavazzi L, Santinelli V. Radiofrequency catheter ablation and antiarrhythmic drug therapy: a prospective, randomized, 4-year follow-up trial: the APAF study. Circ Arrhythm Electrophysiol. 2011;4(6):808-14. Epub 2011/09/29. doi: 10.1161/CIRCEP.111.966408. PubMed PMID: 21946315.

11. Spertus J, Dorian P, Bubien R, Lewis S, Godejohn D, Reynolds MR, Lakkireddy DR, Wimmer AP, Bhandari A, Burk C. Development and validation of the Atrial Fibrillation Effect on QualiTy-of-Life (AFEQT) Questionnaire in patients with atrial fibrillation. Circ Arrhythm Electrophysiol. 2011;4(1):15-25. Epub 2010/12/17. doi: 10.1161/CIRCEP.110.958033. PubMed PMID: 21160035.
12. Wokhlu A, Monahan KH, Hodge DO, Asirvatham SJ, Friedman PA, Munger TM, Bradley DJ, Bluhm CM, Haroldson JM, Packer DL. Long-term quality of life after ablation of atrial fibrillation the impact of recurrence, symptom relief, and placebo effect. J Am Coll Cardiol. 2010;55(21):2308-16. Epub 2010/05/22. doi: 10.1016/j.jacc.2010.01.040. PubMed PMID: 20488300.
13. Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, Cobb FR, Pryor DB. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). Am J Cardiol. 1989;64(10):651-4. Epub 1989/09/15. doi: 0002-9149(89)90496-7 [pii]. PubMed PMID: 2782256.
14. Ware J, Jr., Snow KK, Kosinski M, Gandek B. SF-36 Health Survey: Manual & Interpretation Guide. Boston: The Health Institute, New England Medical Center; 1993.
15. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy. 1990;16(3):199-208. Epub 1990/11/05. PubMed PMID: 10109801.
16. Turpin RS, Ozminkowski RJ, Sharda CE, Collins JJ, Berger ML, Billotti GM, Baase CM, Olson MJ, Nicholson S. Reliability and validity of the Stanford Presenteeism Scale. J Occup Environ Med. 2004;46(11):1123-33. Epub 2004/11/10. PubMed PMID: 15534499.
17. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4(5):353-65. Epub 1993/10/05. PubMed PMID: 10146874.

8. SOFTWARE

SAS® Software Version 9.4 or higher. SAS Institute Inc., Cary, NC, USA.

9. SAP REVISION

Revision Date	Section	Summary of Revision	Reason for Revision
May 5, 2018	6.1.1	100 – [(4 – 4) / 4 x 6] x 100 = 100 – [0 / 36] x 100 = 100 Was changed to: 100 – [(4 – 4) / 4 x 6] x 100 = 100 – [0 / 24] x 100 = 100	The St. Jude scoring instructions included an error in the interpretation example. Although it did not change the scoring instructions, the error was corrected to avoid confusion.
June 1, 2018	6.1.2	Follow-up was corrected from 48 months to 60 months	We had sufficient precision to use data out to 60 months.
June 29, 2018	6.2.1	The Likert responses for the SPS were reversed.	Upon coding the SPS-6, it was noted that the Likert responses were reversed and required correction for accurate scoring.
August 4, 2018	6.1.1	A change to AFEQT response allocation was made as follows: "Patients who haven't had A-Fib in more than a year have problems answering AFEQT questions. Discussed with Drs. Spertus & Mark. If says 'no AF symptoms' use response options : o "Not Bothered at All OR I Did Not Have This Symptom" o "Not At All Limited" o "Not At All Bothered" o "No Difficulty At All" o "Not Applicable" o "Strongly Disagree" (work questions) FOR intervals > 18 months: If patient states " I haven't had Atrial Fibrillation >1 year ago OR say "I was never aware of having atrial fibrillation", use skip pattern: Brief Follow-Up QX (18, 30 and 42 Mo intervals) skip qxs # 3-7 Full Follow-up QX (24, 36, and 48 Mo intervals) skip qxx # 15-19 Proxy QX (18, 24, 30, 36, 42, 48 Mo intervals) skip qxs # 15-17	A change to the AFEQT questionnaire data collection was made on May 12, 2012 to address patient confusion about the questionnaire. This change deviated from the published instructions for scoring the AFEQT, and our analysis was adjusted to adhere to rules set during the conduct of the trial.
August 13, 2018	6.1.1	The AFEQT scoring is to include only responses with at least 50% of responses in each domain.	The 50% rule was not included in the official St. Jude AFEQT Scoring Instructions. However, further review of the validation paper (reference 11) states, "at least 50% of completed responses for each domain are required to calculate a meaningful score" so

			this was added to the scoring code.

955