

# DEFINE AFib Statistical Analysis Plan

Revision [2.0]

Page 1 of 25

Form

Medtronic


## Medtronic Statistical Analysis Plan

Clinical Investigation Plan Title	DEFINE AFib
Clinical Investigation Plan Identifier	MDT20024
Clinical Investigation Plan Version	Comprised of: CIP Version 4.0, 29JUL2022  DEFINE AFib Site Assisted Phase Clinical Addendum Version 2.0, 29JUL2022
Statistical Analysis Plan Version Date	10JAN2023
Sponsor/Local Sponsor	Medtronic, Inc. 8200 Coral Sea Street NE Mounds View, MN U.S.A 55112 Phone: 1-800-328-2518
Lead Principal Investigator(s)	Dr. Jonathan Piccini Associate Professor of Medicine Duke University

### Confidentiality Statement

The information contained in this document is confidential and the proprietary property of Medtronic. Any distribution, copying, or disclosure without the prior written authorization of Medtronic is strictly prohibited. Persons to whom the information is disclosed must know that it is confidential and that it may not be further disclosed by them.

## Table of Contents

1.	Version History .....	3
2.	List of Abbreviations and Definitions of Terms.....	3
3.	Introduction.....	3
4.	Study Objectives .....	5
4.1	Primary Objective.....	5
		
5.	Investigation Plan .....	6
6.	Determination of Sample Size.....	9
7.	Statistical Methods .....	10
7.1	Study Subjects.....	10
7.2	General Methodology.....	10
7.3	Center Pooling.....	11
7.4	Handling of Missing, Unused, and Spurious Data and Dropouts .....	12
7.5	Adjustments for Multiple Comparisons .....	12
7.6	Demographic and Other Baseline Characteristics .....	13
7.7	Treatment Characteristics.....	13
7.8	Interim Analyses.....	13
7.9	Evaluation of Objectives .....	13
7.10	Safety Evaluation.....	23
7.11	Health Outcomes Analyses .....	23
7.12	Changes to Planned Analysis.....	23
8.	Validation Requirements .....	24
9.	References .....	24

## 1. Version History

Version	Summary of Changes	Author(s)/Title
1.0	Not Applicable, New Document	Joao Monteiro, Principal Statistician
2.0	Updated to align to version 4.0 of CIP and to updated SAP template	Jeff Lande, Senior Principal Statistician

## 2. List of Abbreviations and Definitions of Terms

Abbreviation	Definition
AF	Atrial Fibrillation
AI	Artificial Intelligence
AUROC	Area Under Receiver Operating Characteristic
CIP	Clinical Investigational Plan
CRF	Case Report Form
CV	Cardiovascular
DIC	Deviance information criterion
EHR	Electronic Health Record
HCU	Healthcare Utilization
HF	Heart Failure
ICM	Insertable Cardiac Monitor
iOS	iPhone Operating System (Apple's mobile operating system developed and distributed by Apple Inc. required for use of it's devices).
IRB	Institutional Review Board
LMM	Linear mixed model
MDT	Medtronic
ML	Machine Learning
NPV	Negative predictive value
PPV	Positive predictive value
QoL	Quality of Life
RDC	Remote data capture
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SME	Subject- Matter Expert
TIA	Transient Ischemic Attack
WAIC	Widely Applicable Information Criterion / Watanabe-Akaike Information Criterion

## 3. Introduction

This Statistical Analysis Plan (SAP) has been designed to document, before data are analyzed, the rationale for the study design, and the planned analyses that will be included in study reports. This SAP

does not limit the analysis in reports. Additional analysis of the study data beyond this plan may be needed. This SAP is developed for the DEFINE AFib study, based on the Clinical Investigation Plan (CIP) version 4 dated July 29, 2022.

Atrial fibrillation (AF) is the most frequent clinically significant cardiac arrhythmia. When diagnosed clinically, AF is associated with increased risk for stroke, heart failure (HF), cardiovascular (CV) morbidity and mortality. It is also a major contributor to the costs of health care. The prevalence of AF is predicted to rise significantly in the coming years due to the aging of the population and growing frequency of other AF risk factors. Considering the rising burden and clinical sequelae of AF, strategies that enable early detection and appropriate clinical management are needed.

Insertable cardiac monitors (ICMs) now allow for continuous arrhythmia monitoring for up to five years. These devices have enabled early detection of AF that can be asymptomatic, infrequent, and/or of short duration. Accordingly, AF detected by ICMs may not be diagnosed through traditional clinical assessment. Studies have demonstrated significantly higher rates of AF detection with ICM versus conventional monitoring in patients with a recent cryptogenic stroke (Sanna et al. 2014) and in patients undergoing catheter ablation for AF (Kapa et al. 2013). Moreover, four recent trials have observed high rates of AF detection (21%- 40% over 12-30 months of follow-up) with ICM monitoring in patients who have risk factors for AF and stroke, but no clinical history of AF. Two of these studies required conventional cardiac monitoring at baseline. Together these studies demonstrate that continuous monitoring with an ICM can identify AF that would not be readily detected through standard clinical assessment.

Prior studies have demonstrated that device-detected AF is associated with the same clinical outcomes as AF diagnosed through standard clinical practice, including stroke, mortality, and heart failure. However, AF detected on implantable cardiac devices including ICMs can be transient and short in duration. The amount of device-detected AF required to increase the risk of cardiovascular morbidity and mortality is unclear. Prior studies have observed device-detected AF of different thresholds ranging from 6 minutes to 24 hours is associated with an increased risk of stroke. While stroke risk seems to increase with longer durations of device-detected AF, this relationship is not linear, and is impacted by the presence of other stroke risk factors. Even less evidence is available for other adverse consequences of AF such as mortality, healthcare utilization and quality of life. Accordingly, physicians are not clear when or how to intervene when AF is detected by ICM monitoring. A key requirement to advance the management of AF patients is the development of tools that can assess important variables such as AF burden or patterns of AF that, alone or in combination with other clinically important variables, improve the risk prediction for AF-associated clinical outcomes with greater efficacy than current risk stratification modalities. Therefore, the purpose of this study is to evaluate the association between complex patterns of device-detected AF and elevated healthcare utilization, reduced quality of life (QoL) and specific clinical outcomes in patients with a LINQ ICM. These data will improve our understanding of how data from the LINQ ICM can be used to guide the management of patients with AF.

#### 4. Study Objectives

### 4.1 Primary Objective

To evaluate whether summary and episodic measurements collected by market-released LINQ ICMs are able to predict increased AF-related healthcare utilization (HCU)

**Endpoint:** Confirmed healthcare visit in the inpatient hospital, outpatient hospital, clinic/office, emergency department, or other care location (including remote visits) where AF was a reason or suspected reason for healthcare interaction

[illegible]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

## 5. Investigation Plan

DEFINE AFib is an observational, post-market clinical study intended to leverage machine learning to evaluate the association between complex patterns of device-detected AF and elevated healthcare utilization, reduced QoL and specific clinical outcomes in patients with a LINQ ICM. These data will improve the business understanding of how data from the LINQ ICM can be used to guide the management of patients with AF.

This study is expected to be conducted at approximately 30 centers in the United States, and it will enroll approximately 5,000 patients (see section 13.1 of the CIP). Study subjects will be followed for up to 5 years, until study closure, or until their LINQ ICM either completes its service life or is explanted, upon which time they will no longer be followed. Study closure may occur when Medtronic requirements have been satisfied per the CIP and/or following a decision by Medtronic or applicable external parties (i.e. institutional review board (IRB) or regulatory authority) to discontinue the study.



Enrollment is anticipated to last for up to 3 years. The estimated total study duration is up to 7 years, representing a three-year enrollment period and up to 5 years of follow-up for all patients.

## Inclusion Criteria

- Patients with a Reveal LINQ or LINQ II ICM who have a self-reported history of AF and who have received an ICM for the following device logged indications, categorized into the following groups:
  - **Stroke:** Cryptogenic stroke indication
  - **AF management:** AF management and post-ablation management indications
  - **Suspected AF:** Suspected AF and palpitations indications
- Individual access and ability to use an Apple iPhone® compatible with Medtronic's research app (iOS v. 13.X or higher)
- Patient is willing and able to comply with the protocol, including CareLink transmissions (requires adequate connectivity), remotely administered instructions, and remote survey participation
- Patient is 22 years of age or older
- Located in the United States, with CareLink managed through servers located in United States (50 states or District of Columbia)
- Valid email address self-reported at enrollment
- Patient must be able to read and write in English
- ICM device managed by a physician who is part of a participating network/clinic

## Exclusion Criteria

Patients with > 24 months elapsed time from recorded LINQ ICM implant or > 48 months elapsed from recorded LINQ II ICM implant date.

## Data Collected

The following data will be collected passively through Medtronic's research app. Data will be collected on an ongoing basis following documented patient consent, and until study termination or patient withdrawal. Following consent, patients will create a profile, opt in to sharing location, Electronic Health Record (EHR) data and other optional components of the study and the patient will have the opportunity to change data permissions that determine collectable data throughout the study. If patients consent to sharing their EHR data, that patient's EHR data will be pulled via the Apple Health app or through direct access to servers at each patient's site of service when available.

- **Electronic Health Records**
  - Clinical Vitals
  - Diagnoses/Conditions
  - Lab results
  - Medications

- Medical procedure/clinical actions
- Patient mortality
- Health app data (when available, including but not limited to)
  - Steps counts to assess activity
  - Nutrition
  - Sleep
  - Vitals
    - Glucose
    - Blood pressure
    - SPO<sub>2</sub>
  - Body measurements
    - Weight
    - Height
    - Body mass index
- Location services (when available)
  - Anonymized geofencing triggers

Patient surveys will be administered to assess QoL, HCU, changes in medical management, AF-associated symptoms, and medication use (see Table 1 ).

**Table 1: Survey and Profile Update Cadence**

● Survey Category	● Survey	● Survey Type	● Deployment Prerequisites	● Rules	● Expiration
● Medical History	● Patient history	● Static	● eConsent completed	● Only completed once at joining of study, necessary for enrollment	● No expiration
	● Medication history (MDT developed tool)	● Static	● eConsent completed	● Completed at beginning of study, available to update/modify over study timeline	● No expiration
● Quality of Life	● EQ-5D-5L	● Static	● eConsent completed	● Starting at enrollment and then every 3 months (every 90 days +/- 2 days**)	● 30 days post issuance
	● AFEQT	● Static	● eConsent completed	● Starting at enrollment and then monthly (every 30 days +/- 2 days**)	● 14 days post issuance



<ul style="list-style-type: none"> <li>Healthcare Utilization</li> </ul>	<ul style="list-style-type: none"> <li>HCU</li> </ul>	<ul style="list-style-type: none"> <li>Dynamic</li> </ul>	<ul style="list-style-type: none"> <li>Starting 30 days after enrollment or first triggered event</li> </ul>	<ul style="list-style-type: none"> <li>Starting 30 days after enrollment and then monthly (30 days +/-5 days*) (+/-2days**) (For triggered surveys, including healthcare facility geofence and daily AF burden from CareLink, consult CIP for details)</li> <li>HCU survey should always be deployed with Symptom and Symptom Severity Survey</li> </ul>	<ul style="list-style-type: none"> <li>7 days post-issuance</li> </ul>
<ul style="list-style-type: none"> <li>Symptom and Symptom Severity</li> </ul>	<ul style="list-style-type: none"> <li>MAFSI and modified EHRA symptom severity</li> </ul>	<ul style="list-style-type: none"> <li>Dynamic</li> </ul>	<ul style="list-style-type: none"> <li>Starting 30 days after enrollment or first triggered event</li> </ul>	<ul style="list-style-type: none"> <li>Starting 30 days after enrollment and then monthly (30 days +/- 5 day*)(+/-2days**) (For triggered surveys, including healthcare facility geofence and daily AF burden from CareLink, consult CIP for details)</li> <li>Symptom and Symptom Severity survey should always be deployed with HCU survey</li> </ul>	<ul style="list-style-type: none"> <li>7 days post issuance</li> </ul>
<ul style="list-style-type: none"> <li>Medication</li> </ul>	<ul style="list-style-type: none"> <li>SMAQ</li> </ul>	<ul style="list-style-type: none"> <li>Static</li> </ul>	<ul style="list-style-type: none"> <li>Starting 90 days after enrollment (+/-2 days**)</li> </ul>	<ul style="list-style-type: none"> <li>Starting 90 days after enrollment and then every 3 months (every 90 days +/-2days**)</li> </ul>	<ul style="list-style-type: none"> <li>30 days post-issuance</li> </ul>
<ul style="list-style-type: none"> <li>Patient App</li> </ul>	<ul style="list-style-type: none"> <li>Patient impact</li> </ul>	<ul style="list-style-type: none"> <li>Static</li> </ul>	<ul style="list-style-type: none"> <li>Starting 6 months after enrollment (180 days +/- 2 days**),</li> </ul>	<ul style="list-style-type: none"> <li>Starting 6 months after enrollment and then every 6 months (every 180 days +/- 2 days**)</li> </ul>	<ul style="list-style-type: none"> <li>90 days post-issuance</li> </ul>

## 6. Determination of Sample Size

A retrospective analysis of the Optum EHR deidentified database (2007-2019), linked to the Medtronic CareLink database of insertable cardiac monitoring (ICM) devices, suggests that in a cohort of 5,000 patients from the intended population of ICM patients, 600 (12%) will have at least one ischemic stroke event and 2,350 (47%) will have at least one healthcare utilization in a five-year follow-up period. Of the healthcare utilization events, 1,050 (21%) will be AF-related. It is expected there will be approximately 5.2 million follow-up days. Given the large number of follow-ups, the ratio of events to follow-up days is extremely rare. Traditional classification approaches (CART, random forest) require under/oversampling

methods and control parameters for adequate convergence. In the retrospective analysis, the target window was defined as the five days prior to an event (thereby over-sampling events by a factor of five) and then under-sampled controls to a 1:1 ratio with over-sampled cases to create a balanced data set. Setting a minimum node size of 80 patients and requiring the minimum node size to range between 2% and 5% of balanced data, adequately converged to results. The undersampling and model fitting exercise were repeated 100 times to bootstrap an estimate of variable importance for each variable in the model. Using this methodology for convergence and estimation, a minimum of 160 (80/0.05/10) to 400 (80/0.02/10) events are needed for analyzing predictor variable influence and 320 to 800 events for creating a prediction algorithm that requires training, validation, and test partitions as part of its methodology. This latter range is sufficient for deep learning approaches (e.g., LSTM autoencoder, CNN classification network) for predicting extremely rare events.

A minimum of 320 to 800 events implies a 5,000 patient cohort is minimally sufficient for algorithm development. This, however, assumes patients enroll immediately after implant and does not consider attrition. Moreover, this analysis showed that 30% of patients will miss at least 50% of their daily CareLink transmissions. Considering these challenges, it is reasonable to expect as little as 50% of our cohort to provide complete data. In such a scenario, a 5,000 patient cohort would be sufficient for modeling health care utilization (1,175 events, 525 AF-related) but would be insufficient for modeling ischemic stroke or for modeling outcome events by smaller subgroups.

## **7. Statistical Methods**

---

### **7.1 Study Subjects**

#### **7.1.1 Disposition of Subjects**

A STROBE flow diagram will be used to describe the disposition of study subjects for analysis of the primary objective. This flow diagram will keep track of the number of patients for the following categories: enrolled, withdrawn, death, exited, > 90 days without transmitting device data, completed at least 1 survey, completed 75% of all tasks.

#### **7.1.2 Clinical Investigation Plan (CIP) Deviations**

Due to the nature of this study, site level deviations will not be collected, only Study level deviations will be collected by the sponsor. Patients who do not meet inclusion/exclusion criteria will be excluded from analysis.

#### **7.1.3 Analysis Sets**

All patients who enroll and successfully install the study app will be included in the full analysis set. Additional data inclusion/exclusion constraints specific to an objective are specified within the analysis methods.

### **7.2 General Methodology**

Data from all enrolled patients will be summarized, but analysis of each objective will account for the completeness of the data in the context of addressing the objective. In particular, patients contributing to objectives that are modeling HCU-based outcomes should have provided a minimal amount of

information to ensure that HCU data has been made available. This can be accomplished through the completion of monthly and triggered surveys or by sharing EHR information through the study app or via sites. Complete device data should be available for all patients. Patients should have regularly transmitted CareLink data throughout follow-up. If no device data or limited device data is available from a patient, that patient will need to be either excluded from analysis or follow-up will need to be limited to the period of time up until device data is not available.

For patients who enroll at the time of LINQ ICM implant, all follow-up data will be analyzed from the time of implant, if the patient completes all baseline enrollment tasks within 7 days of the LINQ ICM implant. Baseline enrollment tasks include enrollment, signing their consent, and completing all baseline survey questionnaires (patient profile, medical history, medication history, EQ-5D-5L, AFEQT). For patients who are enrolled more than 7 days after the LINQ ICM implant or who completed enrollment tasks more than 7 days post-implant, follow-up will begin when all baseline enrollment tasks have completed.

The final data analysis will be performed by a Medtronic statistician or designee. The timing of the final analysis will be determined during the interim analysis (see Section 7.8). Any change to the data analysis methods described in the protocol will require an amendment only if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the clinical study report.

## 7.3 Center Pooling

The primary objective of this study is to develop risk prediction models and not to compare two arms via a hypothesis testing. Thus, poolability analysis will not be performed in this study. However, the following will be done to reduce the effect of possible heterogeneity across sites:

- When randomizing patients into training/validation/test datasets, the randomization will be stratified by sites
- Predictive models will include site characteristics (e.g. region, size healthcare system) and patient characteristics as input of the model
- Error analysis will be performed to identify cohorts with higher prediction error and diagnose the root cause (e.g. heterogeneity across sites) behind these errors
- CRFs will have the definition of the following events to guide center coordinators to identify them:

**Table 2: Healthcare Utilization Definitions**

HCU	All inpatient hospitalizations, outpatient hospital encounters, emergency department visits, clinic visits, urgent care visits, rehab center or other care location interactions (including remote visits)
-----	--

CV-related HCU	A healthcare utilization relating to the heart and the blood vessels or the circulation (e.g., atrial fibrillation, myocardial infarction, stroke, peripheral vascular disease, heart failure).
AF-related HCU	HCU related to AF, suspected AF, AF treatment, or complications related to AF (including bleeding and drug toxicity as determined by treating clinician).
HF-related HCU	A healthcare utilization related to worsening heart failure signs and symptoms (as determined by treating clinicians) such as (but not limited to) hypervolemic and hypovolemic status requiring the administration, alteration, adjustment, or augmentation of HF therapy (diuretics, inotropes and/or vasodilators etc.) or the utilization of ultrafiltration devices.
Hospitalization	A therapeutic inpatient hospitalization (excludes outpatient and emergency room visits) lasting greater than or equal to 24 hours.
Non-hemorrhagic stroke	Rapid onset of a focal or global neurological deficit or other neurological signs/symptoms consistent with stroke; whereas focal, global cerebral or spinal dysfunction was NOT caused by a non-traumatic intraparenchymal, intraventricular, or subarachnoid hemorrhage, as determined by the treating clinician.
Systemic embolism	An embolus resulting in clinical and objective evidence of sudden loss of end organ perfusion.
TIA	New focal neurological deficit with rapid symptom resolution (usually 1 – 2 hours), always within 24 hours without tissue injury (based on neuroimaging) as determined by the treating clinician.

## 7.4 Handling of Missing, Unused, and Spurious Data and Dropouts

Unless otherwise specified in the evaluation methods, data points with missing values will not be included. The percent of missing data will be tracked of be part of data quality.

## 7.5 Adjustments for Multiple Comparisons

No adjustment for multiple comparisons will be made. This is a single arm study, where the goal in the primary objective is to develop predictive models and not perform hypothesis testing.

## 7.6 Demographic and Other Baseline Characteristics

Descriptive statistics will be used to summarize baseline and demographic characteristics. For categorical variables (e.g. sex), counts and percentages will be employed, while for continuous variables (e.g. age), means, standard deviation, quartiles, minimum, and maximum will be provided. These statistics will be provided for all enrolled subjects and also stratified by LINQ device model.

## 7.7 Treatment Characteristics

This section does not apply since no treatment effect is being measured in this study.

## 7.8 Interim Analyses

An interim analysis was initially planned to be performed when there were observed 500 thousand patient-days OR when it had been 16 months since 1<sup>st</sup> enrollment (whichever comes first). Due to a transition from CIP V3 to CIP V4 (and associated app infrastructure changes) near the time of the planned interim analysis, in addition to slow initial enrollment, the interim analysis timeline was updated to occur within 2 years of the 1<sup>st</sup> enrollment. The objective of this interim analysis will be to re-assess the assumptions and determine if the sample size or follow-up duration of the study needs to be adjusted. Based on the interim analysis, details of how the study objectives will be analyzed (e.g., types of statistical models used) are subject to revision.

## 7.9 Evaluation of Objectives

### 7.9.1 Primary Objective

The primary objective [REDACTED] require development of risk-prediction models for various endpoints using different data source combinations. This section describes a general framework that will be used to develop these models.

#### Prediction models

For each endpoint, two types of prediction modeling will be considered: one focusing on short term predictions and other on long term predictions. For both prediction modelling types, the following machine learning (ML)/artificial intelligence (AI) methods might be considered (but not limited to):

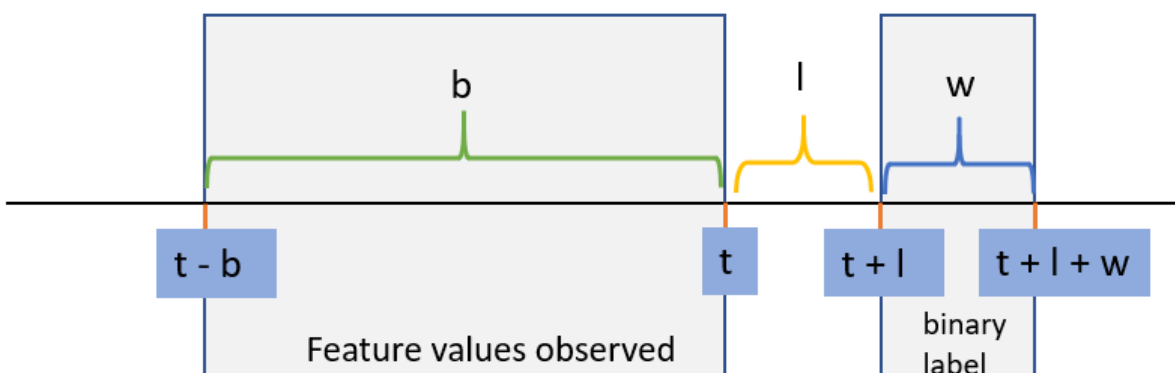
- Tree Boosting
- Temporal Convolutional networks – feedforward neural network (TCN – FFNN) (similar approach presented in Catling & Wolff (2020))

#### Dataset split

Predictive modelling requires the original data to be split into three smaller sets, one for training, one for validation and one for testing. The original data will be randomly split into those three datasets with 70%, 10% and 20% of patients in the training, validation, and testing datasets, respectively. A patient will only contribute data to one of the 3 datasets. To reduce impact of possible heterogeneity across sites, patients and LINQ model, randomization will be stratified by sites, LINQ indication, and LINQ model. Lastly, the testing dataset will only be used after the models under consideration have been trained and validated.

## Data format

All predictive models will be trained using similar labeled examples. In particular, at a prediction time point  $t$  a patient  $i$  generates a labeled example (hereinafter referred as sliding window) which will be composed of the all features considered as relevant to the clinical endpoint (e.g. LINQ device data, EHR data, etc.) collected between  $(t-b)$  and  $t$ , and a binary label indicating if that patient had the clinical event of interest between time points  $t+l$  and  $(t+l+w)$ . Figure 1 illustrates the sliding window concept at time  $t$ . Notice a patient will provide multiple sliding windows.



**Figure 1:** A sliding window at time  $t$  is composed of all relevant features values observed from the most recent  $b$  days and a binary label indicating if event occurred between  $t+l$  and  $t+l+w$ . The rationale of the parameter  $l$  is to give leading time to the physicians to prevent an event for a patient with high risk of having that event between  $t+l$  and  $t+l+w$

## Data transformations

All continuous variables will be centered and scaled. Categorical variables will be converted to into one-hot vectors.

## Handling missing data

Due to the various types of data sources, handling missing data will depend on the requirements of the predictive model and on the data missing patterns observed in the training data. Thus, the decision on



the methods to be used for handling missing data will be left to the discretion of the statistician/data scientist during the model development phase. However, this decision will be made prior of using the testing data.

## Undersampling/Oversampling

Likely the training dataset will have many more instances (i.e. sliding windows) where a clinical event is not observed. Thus, random undersampling/oversampling methods might be needed to build the training dataset. Notice that these approaches will **not** be used to build the validation and test datasets.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]


## Training model

All models under consideration will be training with the goal of minimizing the cross-entropy loss function.

## Tuning

Hyperparameters  $b$ ,  $l$  and  $w$  will be determined via grid search algorithm using the validation set and the Area Under Receiver Operating Characteristic (AUROC) curve as the metric to be maximized. Finite set of reasonable values for  $b$ ,  $l$  and  $w$  will be set during the exploratory data analysis.

The tuning approach for hyperparameters associated with the training model process (e.g. number of epochs, batch size, learning rate, optimizer, etc.) will be left to the discretion of the statistician/data scientist.

## Testing

After all models have been trained and tuned, all final models (i.e. for each clinical endpoint and prediction modeling strategy) will be fitted to the testing dataset resulting in a prediction of each patient's individual probabilities of having the clinical event in the predictive window (i.e. between  $t+l$  and  $t+l+w$ ). More specifically, all models will generate risk prediction (for the clinical event they were designed to predict) for each patient at each day (starting at the  $b^{th}$ ). The true binary values and the corresponding prediction risk scores will be used to compute the area under the receiver operating characteristic (AUROC), sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV). Bootstrapping will be used to compute 95% CIs for these performance metrics.

For each outcome, the performance of the risk prediction models will be presented in contrast to the performance of comparable clinical practices. However, as a guideline, the following scale might be used to help choose the adjective to refer the performance of each model:

- AUROC  $\leq 0.5$ : fail

- $0.5 < \text{AUROC} \leq 0.6$ : poor
- $0.6 < \text{AUROC} \leq 0.7$ : fair
- $0.7 < \text{AUROC} \leq 0.8$ : good
- $0.8 < \text{AUROC} \leq 0.9$ : great
- $\text{AUROC} > 0.9$ : excellent

## Determination of Subjects/Data for Analysis

Subjects in the full analysis set who have completed all enrollment surveys will be included in the primary objective analysis. Subject follow up might be limited by lack of device information.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]

[REDACTED]

- [REDACTED]
- [REDACTED]
- [REDACTED]

Using the training data, for each model (i.e. each distinct random effect structure) under

[REDACTED]

- 1. [REDACTED]
- 2. [REDACTED]
- 3. [REDACTED]

[REDACTED]

- 4. [REDACTED]
- 5. [REDACTED]
- 6. [REDACTED]

[REDACTED]

[REDACTED]

- 7. [REDACTED]
- 8. [REDACTED]

[REDACTED]

[REDACTED]



[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

- [REDACTED]
- [REDACTED]
- [REDACTED]

[REDACTED]

- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]

[REDACTED]

[REDACTED]

- [REDACTED]
- [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

- [REDACTED]
- [REDACTED]
- [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

## 7.10 Safety Evaluation

Summary and listing of any deaths will be reported.

## 7.11 Health Outcomes Analyses

None

## 7.12 Changes to Planned Analysis

This version of the SAP is updated to reflect changes in the CIP from version 3 to version 4. Most of the CIP changes had to do with clarifications around the deployment of patient surveys, which have also been updated in this SAP. Changes in the inclusion criteria and descriptions of study objectives have

also been updated in this SAP. There have been no direct changes in the proposed analyses of study objectives. There was a delay in the timing of the interim analysis, which is discussed in Section 7.8.

## 8. Validation Requirements

---

All programs written to execute the analyses in this SAP will pass through at least Level II validation

## 9. References

---

1. Ariyo, O., Quintero, A., Muñoz, J., Verbeke, G., & Lesaffre, E. (2020). Bayesian model selection in linear mixed models for longitudinal data. *Journal of Applied Statistics*, 47 (5), 890-913.
2. Catling, F.J.R., & Wolff, A. H. (2020). Temporal convolutional networks allows early prediction of events in critical care. *Journal of the American Medical Informatics Association*, 27(3), 355-365
3. Cheng, J., Edwards, L.J., Maldonado-Molina, M., Komro, K. A., & Muller, K. E. (2010). Real longitudinal data analysis for real people: Building a good enough mixed model. *Statistics in Medicine*, 29, 504-520
4. Kapa, S., Epstein, A.E., Callans, D.J., et al (2013). Assessing arrhythmia burden after catheter ablation of atrial fibrillation using an implantable loop recorder: the ABACUS study. *J Cardiovasc Electrophysiol*, 24:875-881.
5. Sanna, T., Diener, H.C., Passman, R.S., et al. (2014) Cryptogenic stroke and underlying atrial fibrillation. *New England Journal of Medicine*, 370:2478-2486.

# DEFINE AFib Statistical Analysis Plan

Revision [2.0]

Page 25 of 25

Form

Medtronic