Statistical Analysis Plan (SAP)
Indication Extension for Medical Devices Using RWE from NESTcc Network Collaborators: Safety and Effectiveness of Cardiac Ablation of Persistent Atrial Fibrillation and Ischemic Ventricular Tachycardia using ThermoCool Catheters
Statistical Analysis Plan Version: 2.0
20 July 2021

Indication Extension for Medical Devices Using RWE from NESTcc Network Collaborators: Safety and Effectiveness of Cardiac Ablation of Persistent Atrial Fibrillation and Ischemic Ventricular Tachycardia using ThermoCool Catheters

Statistical Analysis Plan Version: 2.0

Corresponding to: Protocol Version 1.2

The following individuals have reviewed this version of the Statistical Analysis Plan and are in agreement with the content:

Signature Pages

Guy Cafri
Director, Medical Device Epidemiology and Real World Data Sciences
Johnson & Johnson

Print	Sign	Date
Guy Cafri	DocuSigned by: 0FE42921F1D0423...	7/20/2021

Shumin Zhang
Senior Director, Regulatory RWE & Epidemiology, Medical Device Epidemiology and Real World Data Sciences
Johnson & Johnson

Print	Sign	Date
Shumin Zhang	DocuSigned by: EE3E206DB6D64AA...	7/20/2021

Paul Coplan
Vice President, Medical Device Epidemiology and Real World Data Sciences
Johnson & Johnson

Print	Sign	Date
Paul Coplan	DocuSigned by: B24C1F1FB4E2429...	7/20/2021

Joe Drozda
Director, Outcomes Research
Mercy Health

DocuSigned by:

Joseph P. Drozda, Jr., M.D.	Joseph P. Drozda, Jr., M.D. E339AED0CDB048A...	7/21/2021
Print	Sign	Date

Jiajing Chen
Biostatistician, Outcomes Research
Mercy Health

DocuSigned by:

Jiajing Chen	Jiajing Chen 5E1483BB539E4E1...	7/21/2021
Print	Sign	Date

Guoqian Jiang
Professor, Biomedical Informatics
Mayo Clinic

DocuSigned by:

Guoqian Jiang	D4A7A4754175453...	7/21/2021
Print	Sign	Date

List of Abbreviations

ST	SmartTouch
STSF	SmartTouch SF
AF	Atrial Fibrillation
VT	Ventricular Tachycardia
EHR	Electronic Health Records
AAD	Antiarrhythmic Drugs
BMI	Body Mass Index
CABG	Coronary Artery Bypass Grafting
MI	Myocardial Infarction
PCI	Percutaneous Coronary Intervention

Table of Contents

1. Study Rationale and Design

This study aims to provide evidence using real-world data supporting a label expansion of ThermoCool catheters for patients with persistent atrial fibrillation (AF) or ischemic ventricular tachycardia (VT). Retrospective real-world data from Mercy and Mayo Clinic electronic databases will be used. The first indication expansion being sought is for the ThermoCool Smarttouch® (ST) catheter to treat persistent AF. An investigational device exemption clinical study was conducted for a label expansion for the ThermoCool Smarttouch® SF (STSF) catheter for persistent AF [1] and approved on September 30, 2020 under PMA supplement P030031/S100. This study will use ThermoCool STSF as the comparator to expand the labelled indication for ThermoCool ST among persistent AF patients. The second indication expansion is for the ThermoCool STSF catheter to include an indication for the treatment of ischemic VT (defined by the presence of a history of myocardial infarction (MI) prior to the index ablation procedure in this study). The ThermoCool ST catheter has a label for ischemic VT. Among patients who had procedures performed with ThermoCool STSF, this study will compare the safety outcomes to a performance goal in order to expand the labelled indication among ischemic VT patients.

1.1 Primary Objectives

The primary objectives are:

1) To evaluate the non-inferiority of ablation with the ThermoCool ST catheter relative to the ThermoCool STSF catheter for the treatment of persistent AF on a composite safety endpoint.

2) To evaluate ablation with the ThermoCool STSF catheter relative to a performance goal for the treatment of ischemic VT on a composite safety endpoint.

Determinations about non-inferiority with respect to safety for the first objective (persistent AF) will be based on a hypothesis test of the subject device relative to the comparator device with respect to absolute risk (i.e., risk difference) using data balanced on pre-specified measured covariates. Hypothesis testing for the second objective (ischemic VT) will compare the cumulative incidence of a composite safety endpoint in the subject device group to a performance goal using unadjusted data.

1.2 Exploratory Objectives

The exploratory objectives are:

- To describe the effectiveness of ablation with the ThermoCool ST catheter relative to ThermoCool STSF for the treatment of persistent AF using the propensity score balanced data.

- To describe the effectiveness of ablation with the ThermoCool STSF catheter for the treatment of ischemic VT using the unadjusted data.

- To describe the safety of ablation with the ThermoCool STSF catheter for the treatment of overall VT on a composite safety endpoint using the unadjusted data. The overall VT population will be identified using the same criteria as ischemic VT except removing the requirement of prior MI.

- To describe the safety of ablation with the ThermoCool STSF catheter in a subgroup that excludes non-ischemic cardiomyopathy patients using a composite safety endpoint in the ischemic VT population using the unadjusted data.

- Explore the sensitivity of the safety results for persistent AF to hospital bed size. Remove hospital bed size as a balancing variable for patients with persistent AF in the Mayo data source. Perform the balanced outcome analysis with this modification in Mayo and re-calculate an average difference (all Mercy hospital bed sizes were above the prespecified bed size cut point of 500 so this covariate was not included in the Mercy propensity score model).

- Explore the sensitivity of the safety results for ischemic VT by performing an outcome analysis comparing the ThermoCool STSF group to the ST group using the balanced data.

- Perform exploratory subgroup analyses among patients with a prior prescription of class I or III AAD before the index ablation procedure for the persistent AF population using propensity score balanced data.

Determinations about comparative effectiveness and safety of these exploratory objectives will not be based on hypothesis tests, rather examination of point and confidence interval estimates within each group and differences.

2. Analysis Data Sets

The primary objective will be evaluated on two separate patient populations. The inclusion/exclusion criteria for the **persistent AF population** consists of including patients meeting <u>all</u> the following:

- Underwent an intracardiac catheter ablation for AF by pulmonary vein isolation
- Had a primary or secondary diagnosis of persistent AF associated with the same intracardiac catheter ablation procedure encounter
- The index procedure for a patient is the first recorded intracardiac catheter ablation for AF by pulmonary vein isolation with a primary or secondary diagnosis of persistent AF
- Index procedure performed using the ThermoCool ST catheter or ThermoCool STSF catheter
- At least 18 years of age or older at the time of the index procedure
- At least 6 months of encounter history prior to the index procedure

And excluding patients meeting any of the following:

- Received both the device of interest (ThermoCool ST catheter) and the comparator device (ThermoCool STSF catheter)
- Underwent an intracardiac catheter ablation prior to the index procedure
- Had a surgical cardiac ablation any time prior to or at the same time as the index procedure
- Had a concomitant left atrial appendage occlusion procedure at the time of the index procedure
- Had a concomitant atrioventricular node ablation at the time of the index procedure
- Had a prior heart transplant or long-term heart assist system implantation prior to the index procedure

The inclusion/exclusion criteria for **ischemic VT** consist of including patients meeting all the following:

- Underwent an intracardiac catheter ablation for VT
- Had a primary or secondary diagnosis of VT associated with the same intracardiac catheter ablation procedure
- The index procedure for a patient is the first recorded intracardiac catheter ablation for VT with a primary or secondary diagnosis of VT
- Index procedure performed using the ThermoCool STSF or ThermoCool ST catheter
- At least 18 years of age or older at the time of the index procedure
- At least 6 months of encounter history prior to the index procedure
- Any myocardial infarction diagnosis prior to the index procedure

And excluding patients meeting any of the following:

- Received both the device of interest (ThermoCool STSF) and the comparator device (ThermoCool ST)
- Underwent an intracardiac catheter ablation prior to the index procedure
- Had a surgical cardiac ablation prior to or at the time of the index procedure
- Had a prior heart transplant or long-term heart assist system implantation prior to the index procedure
- Patients with less than 7 days of follow-up due to administrative censoring*

* This censoring event will be present if the latest date of follow-up - date of index procedure < 7 days. The latest date of follow-up is determined by the date that the data is extracted from the EHR.

3. Sample Size Justification

Sample size calculations are based on statistical power for the primary objectives (i.e., safety analyses) proposed in this study. The minimum required sample size was calculated separately for the planned hypothesis tests of non-inferiority of the composite safety endpoint in the persistent AF cohort and ischemic VT cohorts. Below are the details of power calculations.

For persistent AF, hypothesis tests are based on two-group comparison data pooled across data sources using the upper bound of the two-sided 90% Wald confidence interval (section 4.7.3). For this reason, power calculations use a two-sample test for a difference in proportions based on a Pearson chi-square test as implemented in PROC POWER in SAS, version 9.4. For persistent AF, we assume the composite safety endpoint proportion is 0.07 in both arms (i.e., an assumed population difference in proportions of 0), a non-inferiority margin of 0.07, and a one-sided test at the 0.05 alpha-level to determine the minimal N which provided more than 80% power; this sample size was 330 (or 165 per group).

For ischemic VT, hypothesis testing is based on a single group (ThermoCool STSF) with data pooled across the two data sources using the upper bound of the two-sided 90% confidence interval for a one-sample exact test of a proportion (section 4.7.4). For this reason, power calculations use a one-sample exact test for a proportion as implemented in PROC POWER in SAS, version 9.4. For ischemic VT, we assume the composite safety endpoint proportion is 0.075, a performance goal of 0.15, and a one-sided test at the 0.05 alpha-level. Given the sample size of 70 (Table 1), the estimated power is 57%.

Based on the sample sizes available (Table 1), the study is likely adequately powered for persistent AF, but less than adequately powered for ischemic VT.

Table 1. Available Sample Sizes for Each Study Population After Inclusion/Exclusion Criteria Are Applied

	ST (test)	**STSF (comparator)**
Persistent AF		
Mercy	186	763
Mayo Clinic	337	164
Total	523	927
	STSF (test)	**ST (reference)**
Ischemic VT (with prior MI)		
Mercy	35	23
Mayo Clinic	35	84
Total	70	107

4. Analytic Methods to Meet the Primary and Exploratory Objectives

4.1. Treatment

In the persistent AF analysis the subject device group is the ThermoCool ST catheter and the comparison group is the ThermoCool STSF catheter. In the ischemic VT analysis the subject device group is the ThermoCool STSF catheter.

4.2 Covariates

For persistent AF an extensive list of covariates was considered. In 4.2.1 the original list is displayed. However, in some instances there was evidence of a positivity violation (i.e., there are

no observations in one of the device arms for at least one of the levels of a nominal covariate) or a value for a covariate was present for all patients in the data. All modifications to the original covariates for these reasons are described as notes to the tables provided in the Appendix.

4.2.1 Persistent AF Covariates

Covariates ***included in the propensity score model***:

- Calendar year of procedure (2014-2018, 2019-2021)
- Patient Age (<65, 65-74, ≥75 years; based on CHA_2DS_2-VASc categories)
- Patient sex (male, female)
- Race (white, non-white)
- Hospital bed size (<500, ≥500)
- Number of AF ablations performed by cardiac electrophysiologists in prior 12 months (<25, 25-50, >50)
- History of valve replacement (no, yes)
- Mitral valve stenosis (no, yes)
- PCI (no, yes)
- CABG (no, yes)
- BMI* (the latest prior to the index procedure) (continuous variable)
- Anemia (no, yes)
- Hypertension (no, yes)
- Diabetes mellitus (both type 1 and type 2) (no, yes)
- Obstructive sleep apnea (no, yes)
- Vascular disease history (prior MI or peripheral arterial disease or aortic plaque) (no, yes)
- Congestive heart failure (no, yes)
- Chronic pulmonary disease (no, yes)
- Ischemic stroke or hemorrhagic stroke or transient ischemic attack or thromboembolism (no, yes)
- Implantable cardioverter defibrillator or pacemaker (no, yes)
- Hospitalizations: AF related (no, yes)
- Chronic renal disease (stages 3-5) (no, yes)
- Elixhauser comorbidity index (≤3, >3)
- Electrical cardioversion (direct current cardioversion) for AF (no, yes)
- History of supraventricular arrhythmia (no, yes)
- History of ventricular arrhythmia (no, yes)
- Use of class I or III AAD (no, yes)
- Use of class II or IV AAD (beta blocker, calcium channel blocker) (no, yes)
- Use of anticoagulants (no, yes)
- Use of antiplatelets (no, yes)

Note: Demographic information, procedural characteristics, hospital and provider characteristics are recorded at the time of the index procedure. Medications are based on 6 months prior to the index procedure. Medical history, arrhythmia-related information and clinical characteristics use all available information prior to the index procedure unless otherwise specified. * Recorded BMI or calculated BMI based on weight and height (weight in kg/height in m^2), the latest measurement prior to the index ablation procedure. AAD: antiarrhythmic drugs; AF: atrial fibrillation;

BMI: body mass index; CABG: coronary artery bypass grafting; MI: myocardial infarction; PCI: percutaneous coronary intervention

4.3 Endpoint Creation

4.3.1 Composite Safety Endpoint for Persistent AF

The composite safety endpoint for the persistent AF cohort will include several adverse events for patients undergoing ablation procedures. Adverse events evaluated within 7 days of the index procedure include:

- Death
- Acute MI
- Acute stroke/cerebrovascular accident
- Transient ischemic attack
- Thromboembolism
- Heart block
- Pericarditis
- Diaphragmatic paralysis
- Pneumothorax
- Pulmonary edema
- Major vascular access complication or bleeding requiring transfusion

Adverse events evaluated within 30 days of the index procedure include:

- Cardiac tamponade/perforation

Adverse events evaluated within 90 days of the index procedure include:

- Pulmonary vein stenosis
- Atrioesophageal fistula

The composite safety endpoint for persistent AF is a **time-to-event** outcome, given the longer follow-up for some events that make up the composite endpoint. If a patient experienced any of the safety events within the specified time-period for each event, they will be recorded as experiencing the composite safety endpoint. The event time will be the number of days from the index procedure to the date of the *earliest* safety event.

If ***no*** safety events occurred within the specified time-period for each event then patients will be censored and recorded as not experiencing the composite safety endpoint. For these patients censoring time will be measured as the number of days from the index procedure to the date of the *earliest* of three censoring events:

i) Incomplete follow-up due to patient drop-out. This censoring event will be based on absence of a health care encounter documented in the EHR after 7 days of the index procedure. Censoring time corresponding to this censoring event is 7 days.

ii) Incomplete follow-up for an administrative reason. This censoring event will be present if the latest date of follow-up - date of index procedure < 90 days. The latest date of follow-up

is determined by the date the data is extracted from the EHR. Censoring time for this censoring event can range from 0 to 89 days.
iii) End of follow-up period. Censoring time is 90 days.

4.3.2 Composite Safety Endpoint for Ischemic VT
The composite safety endpoint for ischemic VT will include the following adverse events for patients undergoing procedures that occurred within 7 days of the index procedure:

- Death
- Acute MI
- Acute stroke
- Deep venous thrombosis
- Pulmonary embolus
- Complete heart block
- Pericardial effusion with hemodynamic compromise
- Cardiac perforation
- New acute severe mitral or aortic regurgitation
- Arterial dissection
- Vascular injury

The composite safety endpoint for ischemic VT is a binary outcome variable, given the short length of follow-up for all events. If a patient experienced any of the safety events within 7 days, they will be recorded as experiencing the composite safety endpoint. If ***no*** safety events occurred within 7 days then patients will be recorded as not experiencing the composite safety endpoint.

4.3.3 Composite Effectiveness Endpoint for Persistent AF
A composite endpoint will be used in order to evaluate effectiveness for persistent AF over two time periods: through 6 months and through 1 year (after the index date). A 3-month blanking period (healing and stabilization) for persistent AF will be implemented for the assessment of effectiveness outcomes across groups.

The composite endpoint for AF consists of any of the following events after 3 months:
- Rehospitalization for atrial tachyarrhythmia (including AF, atrial tachycardia [AT], and atypical atrial flutter [AFL])
- Rehospitalization for heart failure
- Electrical cardioversion for AF/AT/AFL
- Repeat ablation for AF/AT/AFL

If a patient experienced any of the effectiveness events within the specified time-period, they will be recorded as experiencing the composite effectiveness endpoint. The event time will be the number of days from the index procedure to the date of the *earliest* effectiveness event.

If ***no*** effectiveness events occurred within the specified time-period for each event then patients will be censored and recorded as not experiencing the composite effectiveness endpoint. For

these patients censoring time will be measured as the number of days from the index procedure to the date of the *earliest* of four censoring events:

i) Incomplete follow-up due to patient drop-out. This censoring event will be based on absence of a health care encounter documented in the EHR after 90 days from the index procedure. Censoring time corresponding to this censoring event is 90 days.

ii) Incomplete follow-up for an administrative reason. This censoring event will be present if the latest date of follow-up - date of index procedure < 183 days. The latest date of follow-up is determined by the date that the data is extracted from the EHR. Censoring time for this censoring event can range from 0 to 182 days.

iii) Death. This censoring event will be present if the date of death - date of index procedure < 183 days. Censoring time for this censoring event can range from 0 to 182 days.

iv) End of follow-up period. Censoring time is 183 days.

*Note: When the effectiveness endpoint is calculated for effectiveness at 1 year, 182 and 183 will be substituted with 364 and 365, respectively. Descriptive analysis will be performed to evaluate patients' characteristics with and without at least one-year in-person follow-up data (including both face-to-face visits and remote contact, such as telephone visits) for the exploratory effectiveness evaluation to assess any potential selection bias.

4.3.4 Composite Effectiveness Endpoint for Ischemic VT

A composite endpoint will be used in order to evaluate effectiveness for ischemic VT through 6 months and through 1 year (after the index date).

The composite endpoint for ischemic VT consists of any of the following events after the index procedure:

- Rehospitalization for VT
- Rehospitalization for heart failure
- Repeat ablation for VT

If a patient experienced any of the effectiveness events within the specified time-period, they will be recorded as experiencing the composite effectiveness endpoint. The event time will be the number of days from the index procedure to the date of the *earliest* effectiveness event.

If ***no*** effectiveness events occurred within the specified time-period for each event then patients will be censored and recorded as not experiencing the composite effectiveness endpoint. For these patients censoring time will be measured as the number of days from the index procedure to the date of the *earliest* of four censoring events:

i) Incomplete follow-up due to patient drop-out. This censoring event will be based on absence of a health care encounter documented in the EHR after 7 days from the index procedure. Censoring time corresponding to this censoring event is 7 days.

ii) Incomplete follow-up for an administrative reason. This censoring event will be present if the latest date of follow-up - date of index procedure < 183 days. The latest date of follow-up is determined by the date that the data is extracted from the EHR. Censoring time for this censoring event can range from 0 to 182 days.

iii) Death. This censoring event will be present if the date of death - date of index procedure < 183 days. Censoring time for this censoring event can range from 0 to 182 days.
iv) End of follow-up period. Censoring time is 183 days.

*Note: When the effectiveness endpoint is calculated for effectiveness at 1 year, 182 and 183 will be substituted with 364 and 365, respectively. Descriptive analysis will be performed to evaluate patients' characteristics with and without at least one-year in-person follow-up data (including both face-to-face visits and remote contact, such as telephone visits) for the exploratory effectiveness evaluation to assess any potential selection bias.

4.3.5 Participant Disposition

Disposition of the study participants will be summarized descriptively for the subject categories defined below. Each table below will be populated separately for each data source and pooled across data sources for each indication.

Table 2 (Shell). Disposition of Persistent AF patients for the Safety Endpoint

Disposition	ThermoCool ST	ThermoCool STSF	Total
Patients meeting study inclusion criteria	XXX (100.0%)	XXX (100.0%)	XXX (100.0%)
Patients meeting exclusion criteria*	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Received both the device of interest (ThermoCool ST catheter) and the comparator device (ThermoCool STSF catheter)	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Underwent an intracardiac catheter ablation prior to the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Had a surgical cardiac ablation any time prior to or at the same time as the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Had a concomitant left atrial appendage occlusion procedure at the time of the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Had a concomitant atrioventricular node ablation at the time of the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Had a prior heart transplant or long-term heart assist system implantation prior to the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Incomplete follow-up**	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Death after 7 days from index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Administrative censoring (Data Extract Date-Index Date< 90 days)	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Patient lost to follow-up (absence of a health care encounter documented after 7 days of the index procedure)	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)

Patients experiencing safety events within pre-specified time periods**	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Death[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Acute MI[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Acute stroke/cerebrovascular accident[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Transient ischemic attack[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Thromboembolism[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Heart block[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Pericarditis[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Diaphragmatic paralysis[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Pneumothorax[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Pulmonary edema[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Major vascular access complication or bleeding requiring transfusion[a]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Cardiac tamponade/perforation[b]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Pulmonary vein stenosis[c]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Atrioesophageal fistula[c]	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)

* Percentage is based on patients meeting the inclusion criteria **Percentage is based on patients meeting the eligibility criteria [a] Within 7 days of index procedure [b] Within 30 days of index procedure [c] Within 90 days of index procedure

Table 3 (Shell). Disposition of Ischemic VT patients for the Safety Endpoint

Disposition	ThermoCool STSF	ThermoCool ST	Total
Patients meeting study inclusion criteria	XXX (100.0%)	XXX (100.0%)	XXX (100.0%)
Patients meeting exclusion criteria*	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Received both the device of interest (ThermoCool STSF) and the comparator device (ThermoCool ST)	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Underwent an intracardiac catheter ablation prior to the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Had a surgical cardiac ablation prior to or at the time of the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Had a prior heart transplant or long-term heart assist system implantation prior to the index procedure	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Patients with less than 7 days of follow-up due to administrative censoring	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Patients experiencing safety events within 7 days**	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Death	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Acute MI	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Acute stroke	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Deep venous thrombosis	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Pulmonary embolus	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)

Complete heart block	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Pericardial effusion with hemodynamic compromise	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Cardiac perforation	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
New acute severe mitral or aortic regurgitation	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Arterial dissection	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)
Vascular injury	XXX (XX.X%)	XXX (XX.X%)	XXX (XX.X%)

* Percentage is based on patients meeting the inclusion criteria **Percentage is based on patients meeting the eligibility criteria

4.4 Missing Data

Missing data on the outcomes can occur due to loss to follow-up. These patients will be censored at the time at which follow-up terminated. Missing data on the treatment variable is addressed in the study inclusion criteria. For missing covariate information an informative missing data approach is adopted, such that for nominal variables we create a separate level to indicate a missing value, and for continuous variables the mean is imputed and indicator variable for missing values on that variable is created [2]. The variables with levels representing missing data and missing data indicator variables were included in the propensity score model unless there was a positivity violation (i.e., there are no observations in one of the device arms with missing data on a particular covariate).

4.5 Observational Study Conduct

Covariate balancing methods are used to address the comparisons of a subject device to the comparator device. In our approach, covariate balancing is kept separate from the analysis [3]. A physical separation between the design and analysis is put in place [4-6], such that the person responsible for covariate balancing does not have access to the outcome data. Specifically, one individual created the analytical dataset at each institution (Mercy, Mayo Clinic). Information on the outcome variable was removed from the analytical datasets while retaining a linkage id, and these outcome-removed datasets were provided to an independent statistician who balanced the data. The balanced data will be linked back to the outcome at the home institution in order to perform the outcome analysis (section 4.7). The pooled analyses (sections 4.7.3 & 4.7.4) are relatively straightforward and will be performed by an individual at the Mercy organization.

4.6 Covariate Balancing

The propensity score (PS) will be used for covariate balancing. The methodology described in this section is applied within each data source (Mercy, Mayo Clinic) for both safety and effectiveness endpoints for persistent AF. The methodology will be applied separately within Mercy and Mayo Clinic since this study will use a distributed data network approach where each healthcare system analyzes their own data separately using common methods and treatment effect estimates are pooled subsequently.

4.6.1 Propensity Score Calculation

The method for calculating the propensity score is a multivariable logistic regression model, with K-1 dummy variables for nominal covariates with K levels, and a single variable for each continuous covariate (linear specification).

4.6.2 Method for Covariate Balancing

We considered several PS methods to maximize balance [7] (see appendix, tables 1-4). For ischemic VT, none of the methods considered adequately balanced the data, therefore a comparative design was not pursued in favor of single arm design using a performance goal [8]. For persistent AF a singular method that minimizes imbalance across the two data sources was selected, average treatment effect on the treated weights [9] with weight trimming/winsorizing [10]. Specifically, patients receiving the subject device receive a weight of 1 and those receiving the comparator device (c) receive a weight based on the odds, $w_i = \frac{\hat{e}_i}{1-\hat{e}_i}$, where $\hat{e}_i$= estimated propensity score for patient $i = 1, \dots, n_c$. Furthermore, patient weights in the comparator group only are trimmed/winsorized at the 95th percentile of the weight distribution (i.e., all weights with value above the 95th percentile are set equal to the 95th percentile).

When used with logistic regression to estimate the propensity score, weight trimming has been shown to reduce bias in estimating the treatment effect when the model is mis-specified due to non-linear and non-additive effects, as well as to increase precision of the estimated treatment effects [10]. Use of untrimmed ATT weights has been shown unbiased in simulation studies in estimating the risk difference using survival data [11].

4.6.3. Definitions of Absolute Standardized Differences

Balance is evaluated using absolute standardized differences [12]. When evaluating balance before any covariate balancing for continuous covariates for the subject device (s) and comparator device (c) we use $\hat{d} = \frac{|\bar{x}_S - \bar{x}_C|}{\sqrt{\frac{\hat{v}(\bar{x}_S)+\hat{v}(\bar{x}_C)}{2}}}$ and for binary covariates $\hat{d} = \frac{|\hat{p}_S - \hat{p}_C|}{\sqrt{\frac{\hat{p}_S(1-\hat{p}_S)+\hat{p}_C(1-\hat{p}_C)}{2}}}$.

When using weights, covariate balance for a continuous covariate we replace the means and variance of the comparator group with weighted versions of those estimates [13] $\hat{d}^w = \frac{|\bar{x}_S - \bar{x}_c^w|}{\sqrt{\frac{\hat{v}(\bar{x}_S)+\hat{v}^w(\bar{x}_C)}{2}}}$, where $\bar{x}_c^w = \frac{\sum_{i=1}^{n_c} w_i x_i}{\sum_{i=1}^{n_c} w_i}$, $\hat{v}^w(\bar{x}_c) = \frac{\sum_{i=1}^{n_c} w_i}{(\sum_{i=1}^{n_c} w_i)^2 - \sum_{i=1}^{n_c} w_i^2} \sum_{i=1}^{n_c} w_i (x_i - \bar{x}_c^w)^2$. For a categorical covariate we replace the proportion of the comparator group with a weighted version of that estimate [18] $\hat{d}^w = \frac{|\hat{p}_S - \hat{p}_c^w|}{\sqrt{\frac{\hat{p}_S(1-\hat{p}_S)+\hat{p}_c^w(1-\hat{p}_c^w)}{2}}}$, where $\hat{p}_c^w = \frac{\sum_{i=1}^{n_c} w_i x_i}{\sum_{i=1}^{n_c} w_i}$.

Nominal variables with more than two categories are summarized using generalizations of the above for binary covariates [14].

4.7 Outcome Analysis

For persistent AF, the primary objective is to estimate the covariate balanced risk difference comparing the subject device and the comparator device within each data source and

subsequently averaged across data sources, with the average used for hypothesis testing of non-inferiority for the safety endpoint. Additionally, descriptive analyses for the safety endpoint will be conducted (e.g., unbalanced outcome analysis), as well as exploratory analyses (1.2).

For ischemic VT, the primary objective is to estimate the proportion of patients with safety events who received ThermoCool STSF (unbalanced data), pooled across the two data sources and used to compare against a performance goal for hypothesis testing purposes, following the recommendations of Lu et al [8], since covariate balance was not achieved through propensity score balancing approaches. Descriptive analyses for the safety endpoint will be conducted (e.g., proportion with safety events within catheter group and within data source). Descriptive analyses for the safety endpoint will be conducted in patients who received ThermoCool ST for contextualizing the ThermoCool STSF results. Additional exploratory analyses will also be conducted (1.2).

4.7.1 Method for Estimating a Crude Risk and Risk Difference for Persistent AF

The Kaplan Meier estimate of survival is:

$$\hat{S}(t) = \begin{cases} 1 & \text{If } t < t_1 \\ \prod_{t_j \le t} 1 - \frac{d_j}{Y_j} & \text{If } t_j \le t \end{cases}$$

for times $t_1, t_2 \ldots t_D$ for D distinct times, with d_j events at time j and Y_j patients at risk.

The crude risk is defined for either the subject device group or comparator device group as $\widehat{CI} = 1 - \hat{S}(t)$, with $\hat{S}(t)$ denoting the Kaplan Meier estimate of survival for t=90 days. The variance of the crude risk is represented as $\widehat{\text{var}}\,(\widehat{CI})$, using the Greenwood estimate of the variance. The two-sided 90% Wald confidence interval is: $\widehat{CI} \pm Z_{1-0.05} * \widehat{SE}_{\widehat{CI}}$, where $\widehat{SE}_{\widehat{CI}} = \sqrt{\widehat{\text{var}}\,(\widehat{CI})}$.

The crude risk difference is defined as the difference in the cumulative incidence (CI) between the subject device group and the comparator device group at 90 days. Specifically, the estimated crude risk difference is represented in the subject device group (s; ThermoCool ST) and comparator device group (c; ThermoCool STSF) as $\hat{\Delta} = \widehat{CI}_s - \widehat{CI}_c$, where $\widehat{CI} = 1 - \hat{S}(t)$, with $\hat{S}(t)$ denoting the Kaplan Meier estimate of survival for t=90 days. The variance of the crude risk difference is calculated as $\widehat{\text{var}}(\hat{\Delta}) = \widehat{\text{var}}\,(\widehat{CI}_s) + \widehat{\text{var}}\,(\widehat{CI}_c)$, where each variance term is the Greenwood estimate of the variance. The two-sided 90% Wald confidence interval for the difference is: $\hat{\Delta} \pm Z_{1-0.05} * \widehat{SE}_{\hat{\Delta}}$, where $\widehat{SE}_{\hat{\Delta}} = \sqrt{\widehat{\text{var}}\,(\hat{\Delta})}$. All confidence intervals described in this section are for descriptive reporting of results, not hypothesis testing.

4.7.2 Method for Estimating a Covariate Balanced Risk and Risk Difference for Persistent AF

First, the propensity score is calculated for each patient (section 4.6.1). In the subject device arm, patients receive a weight of 1, therefore the estimate of survival is the conventional Kaplan

Meier estimator described above (section 4.7.1). In the comparator arm, patients receive unequal weights ≠ 1, therefore a weighted Kaplan Meier estimator is used to calculate survival [15]:

$$\hat{S}^w(t) = \begin{cases} 1 & \text{If } t < t_1 \\ \prod_{t_j \le t} 1 - \frac{d_j^w}{Y_j^w} & \text{If } t_j \le t \end{cases}$$

where $w_i = \frac{\hat{e}_i}{1-\hat{e}_i}$, $\hat{e}_i$= estimated propensity score for patient $i = 1, \dots, n_c$, $d_j^w = \sum_{i:T_i=t_j} w_i\, \delta_i$, $Y_j^w = \sum_{i:T_i \ge t_j} w_i$, where T_i is a right-censored event time and δ_i is an event indicator, $\delta_i = 1$ if T_i corresponds to an event and $\delta_i = 0$ if T_i is censored.

A point estimate of the covariate balanced risk difference is calculated as $\hat{\Delta} = \widehat{CI}_s - \widehat{CI}_c$, where $\widehat{CI}_s = 1 - \hat{S}(90)$, $\widehat{CI}_c$= 1- $\hat{S}^w(90)$, with $\hat{S}(90)$ denoting the Kaplan Meier estimate of survival and $\hat{S}^w(90)$ the weighted Kaplan Meier estimate of survival at 90 days. The variance of the covariate balanced risk difference is calculated as $\widehat{var}(\hat{\Delta}) = \widehat{\text{var}}\left(\widehat{CI}_s\right) + \widehat{\text{var}}\left(\widehat{CI}_c\right)$, where $\widehat{\text{var}}\left(\widehat{CI}_s\right)$=$\widehat{var}\left(\hat{S}(90)\right)$ and $\widehat{\text{var}}\left(\widehat{CI}_c\right)$=$\widehat{var}\left(\hat{S}^w(90)\right)$. Estimation of the variance corresponding to the survival estimates in the subject device group, $\widehat{\text{var}}\left(\widehat{CI}_s\right)$, is based on a conventional greenwood estimate because all patients receive a weight of 1. Estimation of the variance corresponding to the survival estimates in the comparator device group, $\widehat{\text{var}}\left(\widehat{CI}_c\right)$, is based on an infinitesimal jackknife [17, 18] because patients do not receive the same weight of 1. The two-sided 90% Wald confidence interval for the difference is: $\hat{\Delta} \pm Z_{1-0.05} * \widehat{SE}_{\hat{\Delta}}$, where $\widehat{SE}_{\hat{\Delta}} = \sqrt{\widehat{\text{var}}\,(\hat{\Delta})}$. The two-sided 90% Wald confidence interval within each catheter group is: $\widehat{CI} \pm Z_{1-0.05} * \widehat{SE}_{\widehat{CI}}$, where $\widehat{SE}_{\widehat{CI}} = \sqrt{\widehat{\text{var}}\,(\widehat{CI})}$. All confidence intervals described in this section are for descriptive reporting of results, not hypothesis testing.

4.7.3 Pooling of Risk Differences Across Data Sources for Persistent AF

The approach to pooling effect estimates (i.e., risk differences) is based on fixed-effect or common-effect model, in which a common effect is assumed across data sources. The choice of a fixed-effect model (vs. a random-effects model) is based on several considerations as highlighted by the American Heart Association: “The choice of a fixed-effect model (ie, inverse variance pooling) should be based on 2 important factors: whether the included studies are functionally identical, meaning they include similar or nearly identical populations, interventions, and methods, and whether the goal of synthesis of results across studies is to compute a common effect size that is applicable to populations similar or identical to those included but not generalizable to other populations.” [19] The types of inferences desired in this context are more consistent with a fixed-effect model, insofar as inferences apply to “*this* collection of studies and say nothing about other studies that may be done later, could have been done earlier, or may have already been done…” [20]. The argument about the nature of inferences is independent of one about effect heterogeneity. An argument for a fixed-effect model based on heterogeneity can be made in this application based on the harmonized approach to study design, in which the inclusion/exclusion criteria, treatment groups, covariates and outcome measurement are the same (or approximately so) across the data sources. In addition, given the relatively small number of studies a fixed-effect model is more compelling because a random-effects model requires a Bayesian approach that assumes a particular underlying effect

size distribution or a frequentist approach with modifications to calculation of the standard error and critical values (based on a t-distribution) that can result in extremely wide confidence intervals in order to obtain reasonable coverage [21, 22].

Calculating the average treatment effect from a fixed-effect model is straightforward [23]. For each data source ($q = 1, 2$) we estimate the risk difference, $\hat{\Delta}_q$ (section 4.7.1). The variance of this estimate is denoted by $\widehat{\text{var}}(\hat{\Delta}_q)$ and a weight is constructed by taking the inverse of this quantity, $w_q = \frac{1}{\widehat{\text{var}}(\hat{\Delta}_q)}$. The average treatment effect across studies is then estimated using a weighted mean, $\widehat{\bar{\Delta}} = \frac{\sum_{q=1}^{2} w_q \hat{\Delta}_q}{\sum_{q=1}^{2} w_q}$. The variance of this weighted mean is then $\widehat{\text{var}}\left(\widehat{\bar{\Delta}}\right) = \frac{1}{\sum_{q=1}^{2} w_q}$ and the standard error is $\widehat{SE}\left(\widehat{\bar{\Delta}}\right) = \sqrt{\widehat{\text{var}}\left(\widehat{\bar{\Delta}}\right)}$. The two-sided 90% Wald confidence intervals is: $\widehat{\bar{\Delta}} \pm Z_{1-0.05} * \widehat{SE}_{\widehat{\bar{\Delta}}}$. This is the only confidence interval used for hypothesis testing for the persistent AF population.

4.7.4 Safety Outcome Analysis in Ischemic VT

The safety outcome analysis among ischemic VT patients is based on an estimate of a simple proportion, $\hat{p} = \frac{\#\, safety\ events}{\#\, patients}$. For proportions calculated for an individual catheter group within a data source, a descriptive summary of the data, $\#\, patients$ is the number of patients in the catheter group and $\#\, safety\ events$ is the number of patients who had a composite safety event in the catheter group. For hypothesis testing, $\#\, patients$ is the total number of patients in the STSF catheter group combined in Mercy and Mayo data and $\#\, safety\ events$ is the number of patients who had a composite safety event in the STSF catheter group combined in Mercy and Mayo data. Confidence intervals (two-sided 90%) and one-sided p-values are based on the Clopper-Pearson exact method.

4.8 Hypothesis Testing

The primary analyses are used to make safety determinations in this study, including: 1) comparison of ThermoCool ST catheter (subject device) relative to ThermoCool STSF (comparator) for the treatment of persistent AF and 2) comparison of ThermoCool STSF catheter (subject device) to a performance goal for the treatment of ischemic VT.

For persistent AF the average risk difference for hypothesis testing is the difference in the cumulative incidence (CI) between the two treatment arms, averaged across the two data sources. The average risk difference is represented as $\bar{\Delta}$. The non-inferiority margin for the risk difference was chosen to be $\delta = 0.07$, corresponding to anticipated cumulative incidence in the ThermoCool STSF group. Therefore, to test the research hypothesis, we consider the following null (H_{01}) and alternative (H_{a1}) hypotheses:

$H_{01}: \bar{\Delta} \geq \delta$

$H_{a1}: \bar{\Delta} < \delta$

The hypothesis test will be based on the upper bound of a two-sided Wald 90% confidence interval for the average across data sources, $\widehat{\bar{\Delta}} \pm Z_{1-0.05} * \widehat{SE}_{\widehat{\bar{\Delta}}}$ (section 4.7.3). Therefore, the hypothesis test is based on a one-sided 5% significance level. A one-sided P-value is calculated as $\Phi(Z), where\ \Phi$ is the cumulative distribution function for a standard normal distribution and $Z = \frac{\widehat{\bar{\Delta}} - \delta}{\widehat{SE}_{\widehat{\bar{\Delta}}}}$

For ischemic VT the safety risk is represented by the proportion p. The performance goal for the risk, PG, is set to 0.15, corresponding to twice the anticipated risk in the ThermoCool ST group. Therefore, to test the research hypothesis, we consider the following null (H_{02}) and alternative (H_{a2}) hypotheses:

$H_{02}: p \geq PG$
$H_{a2}: p < PG$

The hypothesis test will be based on the upper bound of a two-sided Clopper-Pearson exact 90% confidence interval for $\hat{p}$ (section 4.7.4). Therefore, the hypothesis tests are based on a one-sided 5% significance level. A one-sided Clopper-Pearson exact P-value will also be reported.

4.9 Reporting of Results of the Safety Outcome Analyses

Table 4 (Shell). Persistent AF Safety Results for Unbalanced Data

Catheter Group	Cumulative Incidence (90% CI)
	Mercy
ThermoCool ST	X.XXX (X.XXX, X.XXX)
ThermoCool STSF	X.XXX (X.XXX, X.XXX)
Difference	X.XXX (X.XXX, X.XXX)
	Mayo
ThermoCool ST	X.XXX (X.XXX, X.XXX)
ThermoCool STSF	X.XXX (X.XXX, X.XXX)
Difference	X.XXX (X.XXX, X.XXX)

Table 5 (Shell). Persistent AF Safety Results for Covariate Balanced Data

Catheter Group	Cumulative Incidence (90% CI)	Weight
	Mercy	
ThermoCool ST	X.XXX (X.XXX, X.XXX)	-

ThermoCool STSF	X.XXX (X.XXX, X.XXX)	-
Difference	X.XXX (X.XXX, X.XXX)	XXXX.XX
	Mayo	
ThermoCool ST	X.XXX (X.XXX, X.XXX)	-
ThermoCool STSF	X.XXX (X.XXX, X.XXX)	-
Difference	X.XXX (X.XXX, X.XXX)	XXXX.XX
Average Difference	X.XXX (X.XXX, X.XXX)	-
P-value for the Average Difference	X.XXX	

Table 6 (Shell). Ischemic VT Safety Results

Catheter Group	Safety Events	N	Cumulative Incidence (90% CI)
	Mercy		
ThermoCool STSF	XX	XX	X.XXX (X.XXX, X.XXX)
ThermoCool ST	XX	XX	X.XXX (X.XXX, X.XXX)
	Mayo		
ThermoCool STSF	XX	XX	X.XXX (X.XXX, X.XXX)
ThermoCool ST	XX	XX	X.XXX (X.XXX, X.XXX)
ThermoCool STSF Risk for Mayo and Mercy Combined	XX	XX	X.XXX (X.XXX, X.XXX)
P-value for Risk ThermoCool STSF for Mayo and Mercy Combined	X.XXXX		

5. Study Success

The proposed analyses are being conducted in two distinct patient populations, analogous to conducting two separate clinical studies in different patient populations, each to support a separate label expansion. If there is no increased risk of ThermoCool ST relative to ThermoCool STSF on the composite safety endpoint for the persistent AF indication, then the study will be declared a success for that indication. If there is no increased risk of ThermoCool STSF relative to the performance goal on the composite safety endpoint for ischemic VT, then the study will be declared a success for that indication.

Due to the low statistical power (57%) for the ischemic VT hypothesis test in this study we propose to supplement the NEST test case with an additional data source, a comparative study (ThermoCool STSF vs. ST) in the Premier Healthcare Database. An amendment to the study protocol describing the analysis plan and quality of the additional data source will be provided to FDA prior to conducting the Premier study. If the pooled Mercy/Mayo analysis is statistically significant and the comparative study in Premier is statistically significant then STSF meets the criteria for the label extension. However, given the limited power of the pooled Mercy/Mayo we

propose that STSF would also meet the criteria for a label extension if the Premier study is statistically significant and the Mercy/Mayo point estimate of cumulative incidence safety event is $\leq$ 10% (compared to the expected complication rate of ThermoCool ST agreed to in the protocol by FDA of 7.5%). This is equivalent to 7 safety events out of 70 ischemic VT patients and an upper bound of less than 18%.

6. Summary of Revisions to the Original Version of the SAP

Five candidate designs based on the propensity score were evaluated by the independent statistician who had access to data on the catheter group at the index procedure (STSF or ST) and the covariates for each of the four study populations (AF Mayo, AF Mercy, VT with prior MI Mayo, VT with prior MI Mercy), *but no access to the outcome data*. Based on covariate balancing results the following major changes were adopted to the SAP:

1) For the persistent AF patient populations in both Mayo and Mercy, average treatment effect on the treated (ATT) weights will be used, with weights above the 95th percentile of the weight distribution trimmed/winsorized (i.e., set equal to the 95th percentile).
2) Bootstrapping was originally proposed as a method to calculate the variance of the survival estimate within each catheter group for the covariate balanced persistent AF data. However, bootstrapping performed by the independent statistician resulted in extremely large weights even after trimming (e.g., due to positivity violations in the bootstrap samples). For this reason, the infinitesimal jackknife [17,18] will be used instead for variance calculations in the comparator arm.
3) In consideration of inadequate covariate balance achieved for the ischemic VT data, as well as the possibility of zero events in one or more catheter groups in the two data sources, a comparative analysis was not pursued in favor of single arm design using a performance goal, as proposed by Lu et al. [8]. Instead, the primary objective for the ischemic VT patient populations is to evaluate whether patients with ischemic VT, who had a procedure performed with ThermoCool STSF (pooled across Mercy and Mayo), result in safer outcomes than a performance goal for the composite safety endpoint.

References

1. Mansour M, Calkins H, Osorio J, et al. Persistent Atrial Fibrillation Ablation with contact force-sensing catheter: The prospective multicenter PRECEPT trial. *JACC Clin Electrophysiol* 2020; 6: 958-69.
2. Rosenbaum PR. *Observational Studies*. Springer: New York, NY, 2009.
3. Rubin, DB. The design versus the analysis of observational studies of causal effects: parallels with the design of randomized trials. *Stat Med* 2007; 26: 20-36.
4. Cafri G, Paxton EW. Mitigating reporting bias in observational studies using covariate balancing methods. *Observational Studies* 2018*;* 4: 292-296.
5. Yue LQ. Regulatory considerations in the design of comparative observational studies using propensity scores. *J Biopharm Stat* 2012; 22:1272–1279.
6. Li H, Mukhi V, Lu N, Xu Y, Yue LQ. A note on good practice of objective propensity score design for premarket nonrandomized medical device studies with an example. *Stat Biopharm Res*. 2016; 8: 282-286.
7. Harder VS, Stuart EA, Anthony JC. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol Methods*. 2010; 15:234-249.
8. Lu N, Xu Y, Yue LQ. Some considerations on design and analysis plan on a nonrandomized comparative study utilizing propensity score methodology for medical device premarket evaluation. *Statistic in Biopharmaceutical Research.* 2020*;* 12: 155-163.
9. Hirano K, Imbens G, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 2003; 71: 1161–1189.
10. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PLoS One*. 2011; 6:e18174.
11. Austin PC, Schuster T. The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: a simulation study. *Stat Meth Med Res.* 2016; 25: 2214–2237.
12. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res.* 2011; 46: 399-424
13. Austin, PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med.* 2015, 34:3661– 3679.
14. Yang D, Dalton DE. A unified approach to measuring the effect size between two groups using SAS. SAS Global Forum 2012, Paper 335-2012.
15. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005; 24:3089-110.
16. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Stat Med*. 2016; 35:5642-5655.

17. Jaeckel L. *The Infinitesimal Jackknife*. Bell Laboratories Memorandum #MM 72-1215-11, 1972.
18. Efron B, Gong G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *American Statistician* 1983; 37: 36-48.
19. Rao G, Lopez-Jimenez F, Boyd J, D'Amico F, Durant NH, Hlatky MA, Howard G, Kirley K, Masi C, Powell-Wiley TM, Solomonides AE, West CP, Wessel J; American Heart Association Council on Lifestyle and Cardiometabolic Health; Council on Cardiovascular and Stroke Nursing; Council on Cardiovascular Surgery and Anesthesia; Council on Clinical Cardiology; Council on Functional Genomics and Translational Biology; and Stroke Council. Methodological Standards for Meta-Analyses and Qualitative Systematic Reviews of Cardiac Prevention and Treatment Studies: A Scientific Statement From the American Heart Association. *Circulation*. 2017;136:e172-e194.
20. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Meth Meth.* 1998*;* 3: 486–504
21. Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biom J*. 2017; 59:658-671.
22. Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of few small studies in orphan diseases. *Res Synth Methods*. 2017; 8:79-91.
23. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth. Method*. 2010, 1: 97-111.

Appendix

Table 1. Persistent AF: Mercy Balance Results

Variable	Value	STSF	ST	Crude	Strata	Untrimmed	Trimmed 99	Trimmed 95	Trimmed 90
Age (cat)				0.110	0.334	0.031	0.038	0.069	0.083
	<65	349 (45.7%)	75 (40.3%)						
	65-74	310 (40.6%)	83 (44.6%)						
	>=75	104 (13.6%)	28 (15.1%)						
Sex				0.054	0.248	0.008	0.042	0.099	0.108
	Male	544 (71.3%)	128 (68.8%)						
	Female	219 (28.7%)	58 (31.2%)						
Race				0.038	0.203	0.016	0.010	0.018	0.017
	White	732 (95.9%)	177 (95.2%)						
	Non-White	31 (4.1%)	9 (4.8%)						
Elixhauser Index				0.067	0.210	0.036	0.021	0.023	0.024
	<=3	183 (24.0%)	50 (26.9%)						
	>3	580 (76.0%)	136 (73.1%)						
Index Year				1.991	0.627	0.017	0.024	0.059	0.069
	2014-18	184 (24.1%)	174 (93.5%)						
	2019-21	579 (75.9%)	12 (6.5%)						
BMI		32.8 (6.7)	31.8 (5.8)	0.158	0.155	0.031	0.034	0.001	0.009
Hx:PCI		49 (6.4%)	13 (7.0%)	0.023	0.313	0.043	0.025	0.001	0.003
Hx:CABG		67 (8.8%)	13 (7.0%)	0.067	0.182	0.003	0.015	0.013	0.009
Hx:HTN		615 (80.6%)	147 (79.0%)	0.039	0.345	0.057	0.058	0.044	0.045
Hx:Diabetes		198 (26.0%)	38 (20.4%)	0.131	0.362	0.010	0.015	0.009	0.006
Hx:OSA		364 (47.7%)	86 (46.2%)	0.029	0.681	0.017	0.006	0.035	0.040
Hx:CHF		318 (41.7%)	77 (41.4%)	0.006	0.272	0.001	0.000	0.005	0.006
Hx:Chronic Pulmonary Disease		215 (28.2%)	52 (28.0%)	0.005	0.130	0.007	0.007	0.023	0.027
Hx:Stroke		96 (12.6%)	20 (10.8%)	0.057	0.268	0.002	0.014	0.008	0.007

Hx:Chronic Renal Disease		127 (16.6%)	29 (15.6%)	0.029	0.068	0.018	0.007	0.009	0.008
Hx:Anemia		62 (8.1%)	12 (6.5%)	0.064	0.230	0.003	0.004	0.023	0.024
Hx:Valve replacement		27 (3.5%)	8 (4.3%)	0.039	0.171	0.013	0.005	0.013	0.019
Hx:Mitral Valve Stenosis		5 (0.7%)	3 (1.6%)	0.091	0.114	0.043	0.045	0.012	0.019
Hx:Vascular Disease		274 (35.9%)	58 (31.2%)	0.100	0.095	0.029	0.026	0.011	0.008
Hx:AF Admission		315 (41.3%)	89 (47.8%)	0.132	0.290	0.050	0.067	0.122	0.132
Hx:DCCV		543 (71.2%)	136 (73.1%)	0.044	0.116	0.008	0.017	0.017	0.018
Hx:Supraventricular Arrythmia		352 (46.1%)	89 (47.8%)	0.034	0.339	0.035	0.031	0.038	0.042
Hx:Vent Arrhythmia		54 (7.1%)	17 (9.1%)	0.076	0.107	0.066	0.032	0.014	0.009
Tx: ICD Pacemaker		61 (8.0%)	18 (9.7%)	0.059	0.316	0.017	0.026	0.056	0.060
Tx: AAD Class I or III		438 (57.4%)	112 (60.2%)	0.057	0.120	0.040	0.054	0.087	0.094
Tx: AAD Class II or IV		697 (91.3%)	172 (92.5%)	0.041	0.121	0.009	0.001	0.035	0.045
Tx: Anticoagulants		681 (89.3%)	167 (89.8%)	0.017	0.153	0.014	0.005	0.029	0.039
Tx: Antiplatelets		618 (81.0%)	155 (83.3%)	0.061	0.298	0.019	0.007	0.015	0.020
Operator Volume				0.759	0.589	0.245	0.272	0.277	0.276
	<25	247 (32.4%)	9 (4.8%)						
	25-50	49 (6.4%)	18 (9.7%)						
	>50	450 (59.0%)	155 (83.3%)						
	Missing	17 (2.2%)	4 (2.2%)						
Number ASDs > 0.20				2	17	1	1	1	1
Mean ASD				0.15	0.26	0.03	0.03	0.04	0.04

Note: When using stratification, covariate balance for a particular covariate is the average balance across strata ($k = 1, \dots, 5; K = 5$), calculated using $\sum_{k=1}^{K} w_{(k)} \hat{d}_{(k)}$, where $w_{(k)} = \frac{n_{s(k)}}{n_s}$ and n_s is the number of patients with the subject device. Hospital bed size not included in the propensity score model because only hospitals with greater than 500 beds performed catheter ablation for persistent AF. Three patients had missing race in STSF and 0 in ST, and including an additional level would produce a positivity violation, therefore these three patients were placed in the non-white category.

Table 2. Persistent AF: Mayo Balance Results

Variable	Value	STSF	ST	Crude	Strata	Untrimmed	Trimmed 99	Trimmed 95	Trimmed 90
Age (cat)				0.279	1.407	0.505	0.151	0.177	0.177
	<65	68 (41.5%)	186 (55.2%)						
	65-74	71 (43.3%)	114 (33.8%)						
	>=75	25 (15.2%)	37 (11.0%)						
Sex				0.010	0.642	0.287	0.169	0.061	0.061
	Male	118 (72.0%)	244 (72.4%)						
	Female	46 (28.0%)	93 (27.6%)						
Race				0.138	0.189	0.248	0.138	0.005	0.005
	White	162 (98.8%)	326 (96.7%)						
	Non-White	2 (1.2%)	11 (3.3%)						
Elixhauser Index				0.267	0.428	0.006	0.189	0.193	0.193
	<=3	39 (23.8%)	121 (35.9%)						
	>3	125 (76.2%)	216 (64.1%)						
Index Year				0.613	0.603	1.032	0.102	0.115	0.115
	2014-18	42 (25.6%)	183 (54.3%)						
	2019-21	122 (74.4%)	154 (45.7%)						
BMI		32.9 (7.8)	32.3 (6.4)	0.086	0.250	0.024	0.253	0.038	0.038
BMI Missing		1 (0.6%)	11 (3.3%)	0.193	0.205	0.250	0.155	0.028	0.028
Hx:PCI		4 (2.4%)	7 (2.1%)	0.024	0.210	0.203	0.170	0.117	0.117
Hx:CABG		4 (2.4%)	7 (2.1%)	0.024	0.172	0.186	0.032	0.077	0.077
Hx:HTN		116 (70.7%)	199 (59.1%)	0.247	0.890	0.999	0.253	0.199	0.199
Hx:Diabetes		37 (22.6%)	47 (13.9%)	0.224	0.411	0.459	0.197	0.210	0.210
Hx:OSA		75 (45.7%)	136 (40.4%)	0.109	0.931	1.275	0.142	0.064	0.064
Hx:CHF		84 (51.2%)	138 (40.9%)	0.207	0.645	0.341	0.191	0.077	0.077

Hx:Chronic Pulmonary Disease		48 (29.3%)	73 (21.7%)	0.175	0.660	0.746	0.123	0.118	0.118
Hx:Stroke		23 (14.0%)	28 (8.3%)	0.182	1.558	0.107	0.258	0.097	0.097
Hx:Chronic Renal Disease		23 (14.0%)	19 (5.6%)	0.285	1.626	0.016	0.259	0.198	0.198
Hx:Anemia		14 (8.5%)	21 (6.2%)	0.088	0.318	0.338	0.116	0.083	0.083
Hx:Valve replacement		6 (3.7%)	14 (4.2%)	0.026	0.260	0.283	0.173	0.024	0.024
Hx:Mitral Valve Stenosis		1 (0.6%)	5 (1.5%)	0.086	0.199	0.173	0.170	0.163	0.163
Hx:Vascular Disease		64 (39.0%)	178 (52.8%)	0.280	0.882	0.082	0.251	0.213	0.213
Hx:AF Admission		86 (52.4%)	131 (38.9%)	0.275	0.597	0.482	0.322	0.162	0.162
Hx:DCCV		111 (67.7%)	139 (41.2%)	0.551	0.549	0.351	0.377	0.458	0.458
Hx:Supraventricular Arrythmia		105 (64.0%)	188 (55.8%)	0.169	0.604	0.893	0.210	0.101	0.101
Hx:Vent Arrhythmia		24 (14.6%)	38 (11.3%)	0.100	0.525	1.036	0.040	0.004	0.004
Tx: ICD Pacemaker		15 (9.1%)	23 (6.8%)	0.086	0.356	0.372	0.268	0.115	0.115
Tx: AAD Class I or III		93 (56.7%)	155 (46.0%)	0.216	0.683	0.083	0.284	0.173	0.173
Tx: AAD Class II or IV		136 (82.9%)	274 (81.3%)	0.042	0.592	0.657	0.459	0.183	0.183
Tx: Anticoagulants		126 (76.8%)	205 (60.8%)	0.351	0.891	1.074	0.581	0.318	0.318
Tx: Antiplatelets		46 (28.0%)	88 (26.1%)	0.044	0.935	0.715	0.071	0.031	0.031
Operator Volume				0.548	1.506	1.360	0.059	0.257	0.257
	<25	23 (14.0%)	95 (28.2%)						
	25-50	135 (82.3%)	198 (58.8%)						
	>50	6 (3.7%)	44 (13.1%)						
Hospital beds				3.012	0.324	0.339	0.479	1.599	1.599
	>=500	10 (6.1%)	301 (89.3%)						
	<500	154 (93.9%)	36 (10.7%)						
Number ASDs > 0.20				14	28	23	13	6	6
Mean ASD				0.29	0.65	0.48	0.21	0.18	0.18

Note: When using stratification, covariate balance for a particular covariate is the average balance across strata ($k = 1, \dots, 5; K = 5$), calculated using $\sum_{k=1}^{K} w_{(k)} \hat{d}_{(k)}$, where $w_{(k)} = \frac{n_{s(k)}}{n_s}$ and n_s is the number of patients with the subject device.

Table 3. Ischemic VT: Mercy Balance Results

Variable	Value	STSF	ST	Crude	Strata	Untrimmed	Trimmed 99	Trimmed 95	Trimmed 90
Age (cat)				0.581	1.082	1.151	1.151	0.778	0.773
	<65	8 (22.9%)	11 (47.8%)						
	65-74	19 (54.3%)	7 (30.4%)						
	>=75	8 (22.9%)	5 (21.7%)						
Sex				0.372	NA	0.649	0.649	0.649	0.649
	Male	33 (94.3%)	19 (82.6%)						
	Female	2 (5.7%)	4 (17.4%)						
Race				0.080	NA	0.298	0.298	0.296	0.295
	White	34 (97.1%)	22 (95.7%)						
	Non-White	1 (2.9%)	1 (4.3%)						
Index Year				0.687	0.805	0.593	0.593	0.343	0.241
	2014-18	10 (28.6%)	14 (60.9%)						
	2019-21	25 (71.4%)	9 (39.1%)						
BMI		31.2 (6.1)	32.1 (5.5)	0.150	0.055	0.028	0.028	0.262	0.313
Hx:Stroke		7 (20.0%)	3 (13.0%)	0.188	0.926	0.336	0.336	0.225	0.179
Hx:AFib		17 (48.6%)	13 (56.5%)	0.160	0.169	0.323	0.323	0.056	0.028
Hx:HTN		34 (97.1%)	20 (87.0%)	0.383	NA	0.548	0.548	0.548	0.548
Hx:Diabetes		16 (45.7%)	9 (39.1%)	0.134	0.420	0.618	0.618	0.299	0.251
Hx:Anemia		6 (17.1%)	3 (13.0%)	0.115	0.786	0.257	0.257	0.115	0.057
Hx:Valve replacement		1 (2.9%)	1 (4.3%)	0.080	NA	0.301	0.301	0.301	0.301
Hx:VT Admission		24 (68.6%)	12 (52.2%)	0.340	0.690	0.254	0.254	0.099	0.048
Tx: ICD Pacemaker		33 (94.3%)	12 (52.2%)	1.081	0.908	1.108	1.108	0.972	0.915
Tx: AAD Class I or III		34 (97.1%)	17 (73.9%)	0.699	0.777	0.220	0.220	0.093	0.145

Tx: Anticoagulants		29 (82.9%)	16 (69.6%)	0.316	0.092	0.193	0.193	0.002	0.014
Operator Volume				0.571	0.494	0.315	0.315	0.164	0.161
	<10	18 (51.4%)	6 (26.1%)						
	>=10	9 (25.7%)	7 (30.4%)						
	Missing	8 (22.9%)	10 (43.5%)						
Number ASDs > 0.20				9	9	14	14	10	9
Mean ASD				0.37	0.60	0.45	0.45	0.33	0.31

Note: When using stratification, covariate balance for a particular covariate is the average balance across strata ($k = 1, \dots, 5; K = 5$), calculated using $\sum_{k=1}^{K} w_{(k)} \hat{d}_{(k)}$, where $w_{(k)} = \frac{n_{s(k)}}{n_s}$ and n_s is the number of patients with the subject device. Elixhauser not included as a covariate in the propensity score model because for STSF there were 0 instances of patients with scores $\leq$3.

Table 4. Ischemic VT: Mayo Balance Results

Variable	Value	STSF	ST	Crude	Strata	Untrimmed	Trimmed 99	Trimmed 95	Trimmed 90
Age (cat)				0.229	1.581	0.828	0.589	0.108	0.124
	<65	11 (31.4%)	31 (36.9%)						
	65-74	13 (37.1%)	35 (41.7%)						
	>=75	11 (31.4%)	18 (21.4%)						
Sex				0.085	0.542	0.537	0.508	0.141	0.123
	Male	31 (88.6%)	72 (85.7%)						
	Female	4 (11.4%)	12 (14.3%)						
Race				0.010	0.315	0.315	0.288	0.036	0.051
	White	33 (94.3%)	79 (94.0%)						
	Non-White	2 (5.7%)	5 (6.0%)						
Elixhauser Index				0.151	1.886	1.642	0.949	0.078	0.059
	<=3	1 (2.9%)	5 (6.0%)						
	>3	34 (97.1%)	79 (94.0%)						
Index Year				1.793	0.458	0.462	0.332	1.019	1.066
	2014-18	5 (14.3%)	68 (81.0%)						
	2019-21	30 (85.7%)	16 (19.0%)						
BMI		32.0 (7.0)	30.6 (6.7)	0.194	0.046	0.622	0.397	0.010	0.025
Hx:CHF		33 (94.3%)	74 (88.1%)	0.220	1.263	1.373	0.745	0.040	0.051
Hx:Stroke		7 (20.0%)	26 (31.0%)	0.253	0.714	0.116	0.287	0.134	0.124
Hx:AFib		21 (60.0%)	43 (51.2%)	0.178	1.402	0.455	0.001	0.007	0.006
Hx:HTN		30 (85.7%)	73 (86.9%)	0.035	1.672	1.332	0.716	0.023	0.019
Hx:Diabetes		9 (25.7%)	24 (28.6%)	0.064	0.809	0.814	0.759	0.186	0.181
Hx:Anemia		5 (14.3%)	15 (17.9%)	0.097	0.584	0.564	0.501	0.116	0.101
Hx:Valve replacement		4 (11.4%)	10 (11.9%)	0.015	0.509	0.461	0.421	0.044	0.039
Hx:VT Admission		26 (74.3%)	57 (67.9%)	0.142	1.473	0.840	0.355	0.099	0.128

Tx: ICD Pacemaker		31 (88.6%)	58 (69.0%)	0.492	0.757	0.899	0.865	0.421	0.399
Tx: AAD Class I or III		25 (71.4%)	66 (78.6%)	0.166	0.726	0.636	0.568	0.009	0.003
Tx: Anticoagulants		21 (60.0%)	40 (47.6%)	0.250	1.374	0.387	0.060	0.021	0.006
Operator Volume				0.444	1.577	1.037	0.526	0.278	0.283
	<10	16 (45.7%)	21 (25.0%)						
	>=10	19 (54.3%)	63 (75.0%)						
Number ASDs > 0.20				8	17	18	17	4	4
Mean ASD				0.27	0.98	0.72	0.49	0.16	0.17

Note: When using stratification, covariate balance for a particular covariate is the average balance across strata ($k = 1, \dots, 5; K = 5$), calculated using $\sum_{k=1}^{K} w_{(k)} \hat{d}_{(k)}$, where $w_{(k)} = \frac{n_{s(k)}}{n_s}$ and n_s is the number of patients with the subject device. A missing indicator for BMI was not included as a covariate in the propensity score model from the propensity score model because STSF had 0 instances.