

Statistical Analysis Plan

Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-Term Effectiveness
“ADAPTABLE”

Protocol History

Version 1.0	May 1, 2020
Version 1.1	December 7, 2020
Version 1.2	February 9, 2021

Prepared by

Hillary Mulder, MS
Biostatistician III
Duke Clinical Research Institute

Lisa Wruck, PhD
Director, Center for Predictive Medicine
Duke Clinical Research Institute

Michael Pencina, PhD
Professor of Biostatistics
Duke Clinical Research Institute

1	Synopsis	3
2	Study Aims	3
	2.1.1 Aim 1	3
	2.1.2 Aim 2	3
	2.1.3 Aim 3	3
3	Data Sources	3
4	Randomization and Drug Allocation	4
5	Study Endpoints	5
	5.1 Primary Endpoint	5
	5.2 Secondary Endpoints	5
	5.3 Primary Safety Endpoint	5
6	Identification of Endpoints	5
7	Summary of Statistical Methods	6
	7.1 Baseline Demographics and Medical History	6
	7.2 Populations for Analysis	7
	7.3 Primary Effectiveness Comparison	7
	7.4 Primary Safety Comparison	8
	7.5 Censoring Rules	8
	7.6 Power	8
	7.7 Secondary Endpoints	9
	7.8 Prior Treatment Effect	10
	7.9 Handling of Missing Data	10
	7.9.1 Outcome Dates	10
	7.9.2 Event-free Survival Outcomes	11
	7.9.3 Longitudinal Outcome Data	11
	7.10 Planned Sensitivity Analyses	11
	7.11 Subgroup Analyses (Heterogeneity of Treatment Effect)	15
8	Tables and Figures	15
9	References	16

1 Synopsis

To identify the optimal dose of aspirin for secondary prevention in patients with atherosclerotic cardiovascular disease (ASCVD), we conducted a pragmatic clinical trial in which 15,000 patients who are at high risk for ischemic events were randomly assigned in a 1:1 ratio to receive an aspirin dose of 81 mg/day vs. 325 mg/day. Study participants were enrolled over 38 months. Maximum follow-up was 50 months (extended from originally specified 44 months). The primary endpoint is a composite of all-cause death, hospitalization for MI, or hospitalization for stroke. The primary safety endpoint is hospitalization for major bleeding with an associated blood product transfusion.

Additional details about the inclusion and exclusion criterion are contained in the study protocol. Many of the following details are also contained in the study protocol, with some differences in proposed statistical methodology.

2 Study Aims

The major aims of the trial are to test the effect of an aspirin policy on efficacy and safety outcomes, as well as to identify if events captured through electronic health record (EHR) data can be used appropriately in future trials. More specifically, the aims are as follows:

2.1.1 Aim 1

To compare the effectiveness of two daily doses of aspirin (81 mg and 325 mg) in reducing a composite endpoint of all-cause death, hospitalization for nonfatal MI, or hospitalization for nonfatal stroke in high-risk ASCVD patients. Secondary endpoints will be the components of the composite primary endpoint as well as coronary revascularization procedures (percutaneous coronary intervention [PCI] or coronary artery bypass grafting [CABG]) performed during study follow-up. The primary safety endpoint will be hospitalization for major bleeding complications with an associated blood product transfusion.

2.1.2 Aim 2

To compare the effects of aspirin in selected, pre-specified subgroups of patients, including women vs men, older vs younger patients, racial minority patients vs white patients, patients with vs. without diabetes mellitus, patients with vs. without chronic kidney disease (CKD), and patients with vs. without treatment with a concomitant P2Y12 inhibitor at baseline.

2.1.3 Aim 3

To develop, refine, and evaluate the infrastructure for PCORnet to conduct multiple comparative-effectiveness trials in the future. This aim will be accomplished with a “phased-in” approach that will allow for an initial testing of the PCORnet infrastructure followed by adjustments to the trial operational plan to most efficiently accomplish Aims 1 and 2. Also, we will carefully monitor the recruitment and enrollment patterns within and across participating Clinical Data Research Networks (CDRNs) and a Health Plan Research Network (HPRN) and will provide regular feedback reports to each CDRN and HPRN to promote consistent recruitment practices.

Note that the statistical analysis plan for Aim 3 will be described in a separate document.

3 Data Sources

Data sources include the following: electronic health record (EHR) data organized according to the PCORnet Common Data Model (CDM) format, Medicare claims data (CMS), private health insurance claims and patient reported outcomes (PRO) and patient reported data collected via the study web portal. Patient portal data was collected on a continual basis via the web portal.

CDM data was extracted through queries issued by the DCRI data coordinating center for a pre-specified number of times throughout and at the end of the trial. Claims data were collected periodically at a few specified time intervals.

4 Randomization and Drug Allocation

Patients were randomized via a patient portal in a 1:1 ratio to receive 81 mg vs. 325 mg of aspirin in an open-label fashion. Patients were asked to obtain their randomized aspirin dose at their local pharmacies. The randomization scheme was established before the inception of enrollment. Randomized treatment assignment has been obtained from the patient portal data. Compliance with the randomized aspirin dose assignment has been obtained from patient reported data.

Additionally, patients have been randomly assigned to follow-up intervals of every 3 vs 6 months. By embedding a secondary randomization for follow-up time intervals, the best methods for optimizing participant adherence, compliance, and retention in this pragmatic trial will be investigated.

5 Study Endpoints

5.1 Primary Endpoint

The primary endpoint of this study is the composite rate of all-cause mortality, hospitalization for nonfatal MI, or hospitalization for nonfatal stroke. Traditional reporting of potential endpoints by study sites with independent adjudication by a Clinical Events Committee has not been done in this trial given the pragmatic nature of the ADAPTABLE study.

We have conducted an endpoint validation plan to ensure the accuracy of the current endpoint classification methodology when compared with clinical endpoint adjudication processes used in traditional clinical trials. See *ADAPTABLE Validation Plan* for the description of the validation plan, as it has been approved by PCORI. Validation will be summarized in the final report as a separate topic and will not be part of the main trial SAP. The results of the validation will not be incorporated into the primary analyses.

5.2 Secondary Endpoints

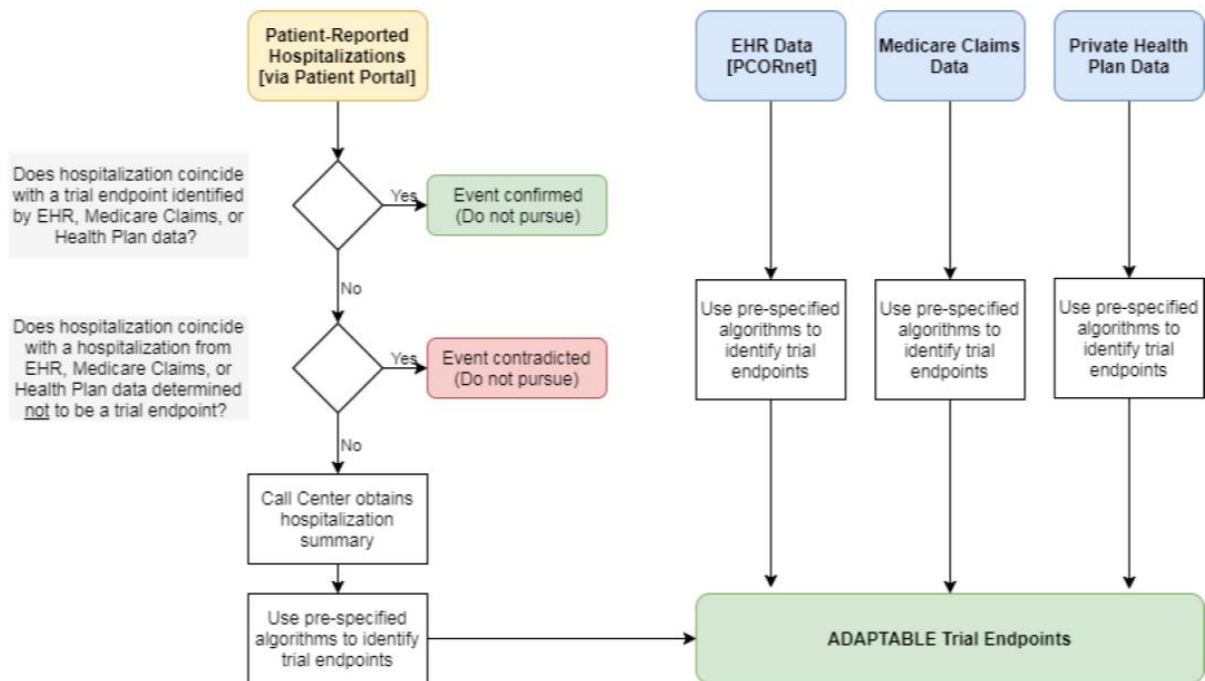
Secondary endpoints include the components of the primary endpoint, coronary revascularization procedures (PCI and CABG), and quality of life and functional status.

5.3 Primary Safety Endpoint

The primary safety endpoint is hospitalization for major bleeding. Bleeding events include intracranial hemorrhage, GI bleeding, bleeding in other locations, or a control of bleeding procedure. To qualify as a major bleeding event, there also must be documented evidence of a blood product transfusion within +/- 7 days of the bleed.

6 Identification of Endpoints

The method for identification of clinical endpoints is documented in the protocol. Endpoints were identified through electronic sources (including EHR [PCORnet CDM], CMS claims or private insurance claims data) or patient reported via the patient portal (see figure below). Confirmation by electronic data is required for patient reported clinical endpoints. Some clinical endpoints, based on patient-reported events that could not be confirmed or refuted by the data sources listed above, were confirmed through medical record review.



For endpoints based on a hospitalization for a specific reason (MI, stroke, bleeding), the primary analysis will include endpoints identified by diagnostic codes both (a) in the principal position, and (b) for which the position is unknown. In the CMS and private insurance claims data, diagnostic code position for inpatient encounters is usually known, and we will use all codes reported. In the PCORnet EHR data, most, but not all, sites consistently use the principal/secondary designation for diagnoses. When diagnosis position is unavailable for a specific code, its position is recorded as “Other/Unknown” in the PCORnet CDM. Therefore, in the PCORnet data, when diagnosis position is available, only codes in the principal position will be used; but when diagnosis position is unavailable, all codes will be used. This requirement will ensure that we do not exclude events from sites that do not record this information. We will use diagnosis codes known to be (and designated as) secondary in sensitivity analyses only. More details on endpoint extraction are contained in the Query Technical Specifications.

Sensitivity analyses, including analyses addressing potential misclassification in endpoint definitions, are described in Section 7.10.

7 Summary of Statistical Methods

7.1 Baseline Demographics and Medical History

Descriptive summaries of baseline demographic and clinical variables will be generated for each randomized treatment arm of the study. Continuous baseline variables will be presented as medians with 25th and 75th percentiles, and discrete variables will be summarized using frequencies and percentages.

Patient age, sex, current smoking status, race and ethnicity are reported in the patient portal data as well as in CDM. These characteristics will be summarized from the patient portal data.

Medical history data obtained from CDM includes: atrial fibrillation, prior coronary revascularization, hypertension, coronary artery disease, congestive heart failure, prior MI, diabetes, COPD/Asthma, chronic kidney disease, peripheral artery disease, prior smoking, significant bleeding disorder, cerebrovascular disease, significant GI bleed, intracranial hemorrhage, and peptic ulcer. Comorbidities are identified through ICD 9 and ICD 10 diagnosis and procedure codes and CPT codes in any care setting at any time during the lookback period. The lookback period is 5 years prior to trial enrollment. Codes can be used from any diagnosis position (principal, secondary or other/unknown).

Medical history information obtained from the CDM is limited to diagnosis and procedures codes associated with encounters within the lookback period that occurred in the recruiting site's healthcare system. There is no expectation that codes associated with older or out-of-system encounters will be captured. This is a standard limitation of EHR data.

7.2 Populations for Analysis

The Intent-To-Treat (ITT) Population will consist of all patients randomized to a treatment group in the study regardless of their compliance with the study medication. For all ITT analyses, participants will be analyzed as randomized.

Patients who were inappropriately randomized due to administrative errors (excluding those taking prohibited medications) were withdrawn from the study prior to initiating study drug. These patients were included in the randomization scheme, but will have no study data beyond that and will not be included in the ITT population.

Given the pragmatic nature of the trial and the expected gaps in data on patient compliance with randomized dose, we will not define the per-protocol (PP) population. Sensitivity analyses are described in section 7.10 that address treatment compliance.

7.3 Primary Effectiveness Comparison

The primary endpoint of this study will be event-free survival from the first event of a composite of all-cause death, hospitalization for nonfatal myocardial infarction, or hospitalization for nonfatal stroke. The primary effectiveness analysis will be performed in the ITT population based on randomized treatment assignment. Event-free survival rates will be compared using Cox proportional-hazards models, equivalent to the log-rank test, using censoring rules described in Section 7.5. The test will be two-tailed and will be performed at an overall α of 0.05. The proportional hazards assumption will be checked for the randomized treatment assignment using weighted Schoenfeld residuals. If there is evidence of non-proportionality, a cautious interpretation of the Cox model hazard ratio will be encouraged, and non-parametric event rate estimates will be emphasized (see below).

In addition to Cox regression, cumulative endpoint event rates will be estimated as a function of follow-up time in each treatment group using Kalbfleisch & Prentice's nonparametric estimator of the cumulative incidence function (CIF). The Kalbfleisch & Prentice CIF estimator is equivalent to the Kaplan-Meier estimator when applied to endpoints that are not subject to competing risks. Cumulative endpoint event rates and differences in cumulative endpoint event rates will be

estimated and presented with 95% confidence intervals. These non-parametric analyses are important for descriptive purposes and will be a focus of interpretation if the event rate curves cross. They may also be used to construct summary measures of treatment effect which are interpretable when the proportional hazards assumption is violated.

Sensitivity analyses for potential misclassification and under-reporting of endpoints and for non-adherence to randomized aspirin dose are described in Section 7.10. An additional sensitivity analysis will be performed to take clustering of patients within sites into account, using the robust sandwich covariance estimator.

7.4 Primary Safety Comparison

The primary safety endpoint of this study will be the first occurrence of hospitalization for major bleeding. The primary safety analysis will be performed by the ITT principle based on randomized treatment assignment. Cumulative endpoint event rates will be estimated using Kalbfleisch & Prentice's nonparametric estimator of the CIF, taking into account the competing risk of all-cause death. Event-free survival rates will be compared using the Fine and Gray method, equivalent to the Cox proportional-hazards model when competing risks are present, using censoring rules defined in Section 7.5. Grey's test is the competing risk equivalent to the log-rank test for comparing the equality of event curves. The test will be two-tailed and will be performed at an overall α of 0.05.

7.5 Censoring Rules

In the event that a patient does not experience an efficacy or safety endpoint, the patient will be censored at the earliest of study end date or maximum follow-up time point at which the patient can be reasonably assumed to be event free. The maximum follow-up time point will be determined from any of the following data sources: CDM censor date, CMS censor date, private health insurance censor date, or the web portal censor date (last point of contact). We expect the study end date to occur on or before all follow-up time points; however, for completeness, censoring dates are defined below. Patients who withdraw consent for trial follow-up will be censored at the time consent is withdrawn. Patients who consent to limited participation follow-up will be censored at the earliest follow-up time point as delineated.

The censoring date for CDM data will be at the site-specific censoring date, which is determined by the PCORnet Distributed Research Network Operations Center based on data curation query results. The censoring date for CMS will be the minimum of the end of enrollment in fee-for-service or the last date claims data are available. The censoring date for private health insurance claims will be the minimum of the end of claims enrollment or the last date claims data are available.

Time to event modeling will include the first occurrence of an event where the time to event is calculated as the event date – randomization date +1.

7.6 Power

Calculations were performed using PASS software. For the primary effectiveness endpoint, power calculations were based on an estimated primary event rate of 5.5% per year (in the higher risk arm), annualized rate of loss to follow-up of 5%, two-sided significance level α of 0.05, 7,500 patients in each treatment arm, enrollment of 38 months and a maximum follow-up period of 44 months. The power of the chosen testing strategy to detect a statistically significant

difference under these assumptions is 85% if the relative risk reduction is 15%, corresponding to a total of 1246 primary effectiveness events.

In November 2019, the DSMB approved extended follow-up in response to updated power calculations based on observed primary event rates. Assuming 7,500 patients in each treatment arm, annualized rate of loss to follow-up of 5%, two-sided significance level α of 0.05, enrollment of 38 months and maximum follow-up period extended to 50 months, the power to detect a relative risk reduction of 15% is 88% when the overall primary event rate is 4.5% in the higher risk arm, corresponding to a total of 1322 primary effectiveness events.

For the primary safety endpoint, power calculations were based on an estimated primary event rate of 2.5% per year (in the higher risk arm), annualized rate of loss to follow-up of 5%, two-sided significance level α of 0.05, 7,500 patients in each treatment arm, enrollment of 38 months and a maximum follow-up period of 44 months. The power of the chosen testing strategy to detect a statistically significant difference under these assumptions is 81% if the relative risk reduction is 20%, corresponding to a total of 642 primary safety events.

7.7 Secondary Endpoints

Secondary endpoints include:

- All cause death
- Hospitalization for nonfatal MI
- Hospitalization for nonfatal stroke
- Coronary revascularization (PCI or CABG)
- Quality of life and functional status – components include:
 - Current Health
 - Physical function
 - Depression
 - Fatigue
 - Sleep disturbance
 - Social roles and activities
 - Pain Interference

Patient reported coronary revascularization events that do not match to one of the electronic sources will not be further verified. As such, these PROs will not be counted as events in the secondary event analyses.

All-cause death will be analyzed using a Cox proportional hazards model. All other time-to-event outcomes will be analyzed using the Fine and Gray method, equivalent to the Cox proportional-hazards model when competing risks are present, with all-cause death as the competing risk. The test will be two-tailed and will be performed at an overall α of 0.05. The proportional hazards assumption will be checked for randomized treatment arm using weighted Schoenfeld residuals. If there is evidence of non-proportionality, a cautious interpretation of the model hazard ratio will be encouraged, and non-parametric event rate estimates will be emphasized, as described for the primary effectiveness analysis.

Quality of life and functional status data are collected on numerical scales (1:excellent health-5:worst health). They will be analyzed as continuous variables. Mixed models using restricted maximum likelihood estimation (REML) will be employed to model trajectory of measures over time by treatment group. Mixed models account for the correlation structure imposed by repeated measures within participants while using all available data, from baseline to the end of the study, regardless of exact follow-up time. The intercept and slope will be modeled as random effects. The covariance structure for random effects will be modeled using an

unstructured form. Time from baseline measurement and randomized treatment arm will be included in the model as fixed effect. Time will be tested for linearity using natural cubic splines. An interaction between randomized treatment arm and time will be assessed, as will the overall effect of randomized treatment arm.

Since these measures are discrete scores, it is expected that normality assumptions will not hold. Natural log and other appropriate transformations will be considered, as well as several distributions and link functions.

No formal adjustments for multiple testing will be performed. All main and sensitivity analysis results will be presented and left to the interpretation of the reader.

7.8 Prior Treatment Effect

We expect that most patients recruited into the ADAPTABLE trial will already be treated with pre-randomization aspirin therapy, given the nature of the inclusion criteria and the known high utilization of aspirin for the secondary prevention of coronary artery disease in the United States. Therefore, a sensitivity analysis for the primary efficacy and safety endpoints with a 10-day landmark will be performed that excludes events occurring in the first 10 days following randomization to account for the expected time period of washout from the pre-randomization aspirin dose.

7.9 Handling of Missing Data

Concerted effort will be made to eliminate or minimize the occurrence of missing data. Participants will grant access to their electronic medical records at enrollment and during study follow-up as well as to have their information searched in national databases. During the course of the trial, missing data has been monitored by the operations team via aggregate reports. The DCRI call center has been responsible for locating participants with 2 consecutive missed study visits who have not been confirmed dead through electronic data queries or through contacts with the site research teams. This process has changed since study initiation so that participants are now contacted if there is no completed visit in the prior 6 months.

If, despite these efforts, missing data occur, we will employ statistical techniques appropriate to the type of data, as described below.

Reasons for missing data will be collected and described, including withdrawal of consent for any follow-up and loss to follow-up for portal visits. Descriptive statistics for key baseline characteristics and clinical events prior to study withdrawal or completion will be presented by subgroups defined by availability of follow-up data and by treatment group. All participants in the defined study population will be accounted for in all analyses and presentations.

7.9.1 Outcome Dates

Any partial or completely missing date for a confirmed primary effectiveness or safety outcome at the time of database lock will be imputed as follows:

- If the day is missing, 15th of the month, or the randomization date (if patient randomized after 15th of the same month and same year) will be used;
- If the month is missing, June, or the randomization month (if patient randomized after June and year of the event is same as randomization year) will be used;
- If the complete date is missing, the midpoint between the date of last known event-free visit and end of follow-up will be used.

7.9.2 Event-free Survival Outcomes

For the primary and secondary analyses based on event-free survival, most participants are expected to have at least one source of data for these endpoints (patient portal with confirmation of reported events, CDM, CMS or private claims). Issues with potential incomplete data will be addressed through sensitivity analyses described in Section 7.10. Participants discontinuing the study prematurely and withdrawing consent for trial follow-up will be considered truly missing and will be censored at the time of discontinuation. This approach might lead to biased results if the mechanism of discontinuation is non-ignorable, i.e. the hazard for a censored participant is not the same as that of uncensored participants, conditional on observed data. If more than 5% of participants withdraw consent for trial follow-up, then a tipping point analysis similar to that described by Little et al (2016) will be conducted to assess impact of potential non-ignorable censoring on inference for the primary analysis. A Weibull model of the primary endpoint will be fit to the entire ITT population, with independent variables including an indicator of withdrawal of consent for electronic follow-up and selected baseline characteristics. Other independent variables may be considered. The Weibull model will yield an estimated hazard at the time of withdrawal of consent for each participant adjusted for selected covariates. We then assume that the hazard for a participant who withdraws consent is different from those who do not and allow that difference to vary between treatment groups. An inflation factor will be applied to the hazards, with different inflation factors applied to each treatment group. The resulting hazards will be used to impute events to the end of trial follow-up assuming a Weibull distribution. A Cox model as specified for the primary analysis will be fit to the resulting dataset and treatment effect hazard ratio will be estimated, with standard errors adjusted using standard multiple imputation combining rules. The resulting inference will be examined across a range of clinically plausible inflation factors; if inference from primary analysis is maintained then the results will be considered robust. To aid in interpretation, the mean number of imputed events will be reported for inflation factors.

7.9.3 Longitudinal Outcome Data

Quality of life and functional status secondary endpoints will be measured longitudinally by patient self-report through the portal or call center. Missing data may occur with missed contacts, participant withdrawal, or refusal to answer questions. The mixed model approach planned for analysis of these endpoints yields unbiased inference in the presence of a missing at random (MAR) missingness mechanism, meaning that the distribution of missing data is the same as that of non-missing data, conditional on observed covariates. As part of planned secondary analyses following the primary final report, we will examine patterns of missingness and characterize participants by missingness pattern and treatment group to assess plausibility of the MAR assumption. If concerns are noted, then a pattern mixture model approach will be considered. If there are no concerns regarding the MAR assumption, but there is significant missing data (more than 10% of participants with more than two missed measures), an inverse probability weighting approach will be considered.

7.10 Planned Sensitivity Analyses

Sensitivity analyses will be conducted to assess robustness of trial conclusions to 1) non-adherence to randomized aspirin dose and 2) under-reporting and misclassification of endpoint data. Sensitivity analyses will be focused on primary analysis of the primary efficacy and safety endpoints, but the methods may be generalized to key secondary endpoints.

Non-adherence. A sensitivity analysis will be conducted in the ITT population in which participant-reported aspirin dose is added to the primary analysis Cox model as a time-varying covariate.

COVID-19. A sensitivity analysis will be conducted in the ITT population in which censoring for the primary efficacy endpoint will occur on December 31, 2019 the original study end date and prior to COVID entering the United States. Although we do not expect a differential impact of randomized aspirin dose, COVID-19 could drastically increase the observed primary efficacy endpoint.

Bayesian Analysis. As part of planned secondary analyses following the primary final report, a sensitivity analysis for the primary efficacy endpoint will be performed using a Bayesian proportional hazards model or a Bayesian piece-wise exponential non-proportional hazards model, pending assessment of the proportional hazards assumption. Using the Bayesian posterior distribution, we will calculate the posterior probability that the unknown treatment specific hazard ratio exceeds thresholds of 1.0, 1.05, 1.10, 1.15 and 1.20 and reciprocals of those numbers. Because Bayesian inferences may be sensitive to the choice of prior distribution, further sensitivity analyses will be performed and reported for a range of possible prior distributions.

Proportional hazards assumption holds.

Bayesian analysis requires the specification of a prior probability distribution representing prior information about the set of unknown model parameters before observing the study data. Because prior information about the treatment effect is limited, we will pre-specify a flat (uniform) prior for the Cox model regression coefficient. The selection of a uniform prior reflects the subjective assessment that prior information about the direction and magnitude of the treatment effect is neutral. It also allows posterior inferences to be dominated by the current study data as opposed to the prior.

Bayesian estimation of the Cox model will be based on the method of Kalbfleisch (1978) which uses the Cox partial likelihood function in place of the full data likelihood. Using the Cox partial likelihood is advantageous because it lends itself to efficient Bayesian MCMC computation via Gibbs sampling and avoids the requirement to specify a prior distribution for the baseline hazard function. Using the partial likelihood in place of the full likelihood has been justified on the grounds that it closely approximates a Bayesian analysis of the full data likelihood when the prior distribution for the baseline hazard function is a highly diffuse gamma process prior (Sinha, Singh, Ibrahim, 2003).

Posterior means and other summaries of the posterior distribution will be calculated using Markov Chain Monte Carlo (MCMC) simulations as implemented in the SAS PHREG procedure. To reduce Monte Carlo error and ensure convergence, we will generate 50,000 sets of simulated parameter values after an initial burn-in period of 2,000 iterations.

Proportional hazards assumption is violated.

The piece-wise exponential model lends itself to simple and efficient Bayesian MCMC computation, allowing the model to be flexible enough to accommodate a variety of shapes for the unknown treatment-specific hazard function. To implement this approach, follow-up time will be divided into discrete time intervals. The treatment-specific hazard function for the primary endpoint will be approximated as a constant function within each treatment group and time interval. A summary measure of treatment effect will then be computed by estimating the ratio of

the cumulative average hazard ratio over follow-up. The cumulative average hazard ratio reduces to the hazard ratio statistic when the proportional hazards assumption holds, but does not rely on the PH assumption for its validity or interpretability. A flat prior will be chosen for all model parameters requiring a prior distribution.

Under-reporting and misclassification of endpoints. The primary and key secondary endpoints are derived from multiple sources reflecting varying levels of data completeness and sensitivity/specificity of electronic phenotype definitions for true diagnosis. It is therefore critical to clarify the assumptions underlying inference around the treatment effect and perform sensitivity analyses to assess robustness of this inference to potential violations of the assumptions.

For the primary analysis (and related secondary analyses), we make the following assumptions:

- Death is reported without error.
- Other components of the primary endpoint may be subject to under-reporting and misclassification but these errors are non-differential, i.e. probability of an error is consistent across randomized treatment groups.
- Patient self-report with confirmation is assumed to have nearly perfect specificity since incorrect self-reports will be excluded by examination of medical record. There may be some under-reporting since identification of the event depends on self-report. Some self-reported events may not be confirmed or contradicted if participant did not consent to confirmation process.
- CDM endpoints may be under-reported since participants may be treated for events outside the CDM health systems or events within the health system may be missed. Misclassification may occur due to imperfect electronic phenotype definitions.
- Claims data (CMS and private) is assumed to be complete for enrolled participants and changes in enrollment status are independent of trial outcome. Misclassification may occur due to imperfect electronic phenotype definitions.

We will perform two sets of sensitivity analyses: 1) to assess impact of under-reporting on the primary analysis and 2) to assess impact of misclassification on the primary analysis. Because available data sources vary across participants, we summarize the potential for error due to under-reporting and misclassification according to each potential combination of available data. To simplify classification of participants, we will not attempt to account for changes in data sources over time, but rather classify participants according to data sources available for more than half of follow-up time.

Scenario	Data sources available				Potential error	
	Self-report	Out of Network Confirmation	CDM	Claims (CMS or private)	Under-reporting	Misclassification
1	X				Yes	Yes
2	X	X			Yes	No
3	X		X		Yes	Yes
4	X			X	No	Yes
5	X		X	X	No	Yes
6			X		Yes	Yes
7				X	No	Yes

8			X	X	No	Yes
9*					Yes	No

*Only information is vital status check at end of trial.

To assess impact of under-reporting due to participants moving or seeking care outside of the CDM health system, we will repeat the primary analysis modifying the censoring rules for CDM data such that the censoring date will be the last date the patient encountered the health system. This will include inpatient and outpatient visits, labs, vitals and prescription medication fills. All other censoring dates will remain the same.

Under the assumption that under-reporting is non-differential between treatment groups, the impact on primary analysis inference would be a loss of power. If the primary result is non-significant then we will assess overall impact of under-reporting, by performing a tipping point analysis to identify how many events would have to be missed to achieve significance. We will take an approach that is adapted and simplified from that described by Little et al (2016) for a sensitivity analysis for missing data. A Weibull model of the primary endpoint will be fit to the entire ITT population, with independent variables including an indicator that participant is at risk for under-reporting (Scenarios 1, 2, 3 and 6 in the table) and selected baseline characteristics. Other independent variables may be considered. Events will be sequentially added for those participants at risk of under-reporting across treatment groups. For each increment, i events will be imputed from the Weibull model 1000 times. A Cox model as specified for the primary analysis will be fit to the resulting dataset and treatment effect hazard ratio will be estimated, with standard errors adjusted using standard multiple imputation combining rules. The increment will be increased until the upper limit of the 95% CI for the HR of the treatment effect crosses 1.0.

To assess impact of misclassification of electronic phenotype definitions, we will repeat the primary analysis modifying the code-based definitions of primary endpoints to include codes in 1) principal position only and 2) any position.

To further assess impact of misclassification, we will perform a tipping point analysis expanding on methods developed by Liublinska and Rubin (2014) for missing data in clinical trials. Positive predictive value and negative predictive value will be varied from 0 to 1 to reclassify events or non-events for those participants at risk of misclassification (Scenarios 1 and 3-8). For each combination of PPV and NPV, the corresponding number of reclassified events will be imputed 1000 times. Reclassified events will randomly selected and set to non-events. Reclassified non-events will be generated from the Weibull model described in tipping analysis for under-reporting. A Cox model as specified for the primary analysis will be fit to the resulting dataset and treatment effect hazard ratio will be estimated, with standard errors adjusted using standard multiple imputation combining rules. A 95% confidence interval will be calculated for the HR of the treatment effect. Resulting inference based on comparing confidence limits to 1.0 will be presented in a heat map with positive predictive value and negative predictive value reported on the x- and y-axes. Note that this analysis does not relax the assumption of consistency of misclassification across treatment groups.

7.11 Subgroup Analyses (Heterogeneity of Treatment Effect)

Subgroup analyses for the primary effectiveness and safety endpoints will be performed on the ITT population in order to explore whether the treatment effect is consistent across subgroups. Subgroup analyses to evaluate variation in treatment effect will be performed on the basis of tests for interaction using the Cox proportional-hazards model with terms for treatment group, the subgroup variable and treatment by subgroup variable interaction. Additionally, treatment effects within each categorical subgroup will be examined separately using Cox proportional-hazards models. Event rates by treatment and HRs with 95% confidence intervals will be reported for each subgroup. Forest plots will be generated displaying the estimated hazard ratios and 95% confidence intervals for each subgroup. For continuous variables, the linearity assumption will be checked for violations using natural cubic splines. In the event that major violations are found, natural cubic splines will be used in the final model. This analysis will be considered primary, but for display purposes, these variables will also be categorized.

The following variables determined at baseline will be examined:

- Age (continuous)
- Race categories (White, Black, and Asian; Hispanic ethnicity)
- Diabetes mellitus
- Chronic kidney disease (serum creatinine > 1.5 mg/dL)
- P2Y12 inhibitor use
- Female sex

We expect homogeneity of treatment effect across subgroups following the results observed in the full sample. Thus, testing for differences between treatment arms within subgroups will be considered exploratory and no claims of heterogeneity will be made based on tests within subgroups.

As a sensitivity analysis, the interaction between randomized treatment arm and internet vs non-internet enrollment will be assessed for the primary efficacy and safety outcomes.

8 Tables and Figures

See appendices for Table Shells.

9 References

Little R.J., D. R. (2012). The Prevention and Treatment of Missing Data in Clinical Trials. *N Engl J Med*, 367:1355-1360.

RJ Little et al. The treatment of missing data in a large cardiovascular clinical outcomes study. *Clinical Trials* 2016

Liublinska V, and Rubin DB (2014). Sensitivity Analysis for a Partially Missing Binary Outcome in a Two-Arm Randomized Clinical Trial. *Stat Med*, 33(24):4170-4185.

Kalbfleisch, J.D. and R.L. Prentice, The statistical analysis of failure time data, second edition. Vol. 360. 2011: John Wiley & Sons.

Kaplan, E. and P. Meier, Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 1958. 53: p. 457-481.

Hernan, M.A., The Hazards of Hazard Ratios. *Epidemiology*, 2010. 21(1): p. 13-15.

Royston, P. and M.K. Parmar, The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, 2011. 30(19): p. 2409-2421.

Péron J., P. Roy, B. Ozenne, L. Roche, M. Buyse. The net chance of a longer survival as a patient-oriented measure of treatment benefit in randomized clinical trials. *JAMA oncology*. 2016 Jul 1;2(7):901-5.