Insulin Resistance Intervention After Stroke Trial

IRIS

Statistical Analysis Plan

Version 1.3

3/1/2015

Table of Contents

Statistical Analysis Plan

Insulin Resistance Intervention after Stroke Trial

Background

Among 400,000 Americans who survive an ischemic stroke or TIA each year, a major source of morbidity and mortality is recurrent stroke and myocardial infarction. Within 5 years of the initial event, 18% of patients will have a recurrent stroke, 9% will have a myocardial infarction, and 6% will die from one of these two conditions. Prevention of further vascular events, therefore, is of major importance in reducing the morbidity and mortality of stroke. Current strategies for preventing vascular events after stroke include antiplatelet therapy, hypertension control, anticoagulation, lipid management, and carotid endarterectomy. Despite these effective strategies, many suitably treated stroke patients still experience recurrent vascular events. New therapeutic approaches are urgently needed.

The Insulin Resistance Intervention after Stroke (IRIS) trial proposes an innovative therapy based on 20 years of accumulating evidence linking insulin resistance to increased risk for stroke and other vascular diseases. The aim of the IRIS trial is to determine if pioglitazone, an agent that reduces insulin resistance and has other potent vasoprotective effects, is effective in lowering the risk for stroke or myocardial infarction among non-diabetic men and women 40 years of age and older with a recent ischemic stroke or TIA and insulin resistance. The IRIS trial will determine the effectiveness of a new strategy, the treatment of insulin resistance in non-diabetic subjects, for preventing recurrent stroke and myocardial infarction after ischemic stroke or TIA. Since insulin resistance is estimated to affect 50% of stroke patients, this innovative treatment has the potential to benefit a large number of patients. The study will also yield important data on the effectiveness of pioglitazone in preventing progression to overt diabetes, and in reducing the risks for all-cause mortality and cognitive decline.

Primary Objective

1. To determine if pioglitazone, compared to placebo, will reduce the overall risk for fatal or non-fatal stroke or fatal or non-fatal MI among non-diabetic men and women 40 years or older with insulin resistance and a recent ischemic stroke or TIA.

   Among non-diabetics with insulin resistance, we hypothesize that pioglitazone will reduce the occurrence of any primary endpoint (fatal or non-fatal stroke or MI) within four years from 27% to 22%.

Secondary Objectives

1. To determine if pioglitazone, compared to placebo, will reduce the risk for recurrent stroke.

   We hypothesize that pioglitazone will reduce the occurrence of recurrent fatal or non-fatal stroke as a discrete outcome.

2. To determine if pioglitazone, compared to placebo, will reduce the risk for acute coronary syndromes (ACS) (acute MI or unstable angina).

   Unstable angina is an important clinical event because it identifies individuals at high risk for MI who need urgent diagnostic and therapeutic intervention. We hypothesize that pioglitazone will reduce the overall occurrence of acute coronary syndromes.
3. To determine if pioglitazone, compared to placebo, will reduce the risk for all-cause mortality.

   We hypothesize that pioglitazone will reduce all-cause mortality because of its vasoprotective effects.
4. To determine if pioglitazone, compared to placebo, will reduce the risk for recurrent stroke, MI, or serious congestive heart failure (CHF).

   We hypothesize that pioglitazone will reduce the occurrence of recurrent fatal or non-fatal stroke, MI or CHF.
5. To determine if pioglitazone, compared to placebo, is effective in preventing progression to overt diabetes.

   Insulin resistance is the principal risk factor for type II diabetes. We hypothesize that pioglitazone, by sensitizing cells to insulin's action, will prevent progression to diabetes.
6. To determine if pioglitazone, compared to placebo, will reduce the risk for cognitive decline.

   We hypothesize that pioglitazone will reduce the risk for cognitive decline by reducing the incidence of clinically apparent and unapparent stroke.

Study Design

The Insulin Resistance Intervention after Stroke Trial (IRIS) is a double-blind, placebo-controlled, multi-center clinical trial in which approximately 4000 non-diabetic patients ≥ 40 years of age with insulin resistance and a recent non-disabling ischemic stroke or TIA are randomly assigned to pioglitazone, a thiazolidinedione (TZD), or placebo. Randomization is stratified by participating site using a random permuted block design of varying block size (4, 6, 8) and treatment allocation ratio of 1:1 between placebo and active drug. Maximum follow-up will be 5 years (plus 21 days allowed after the 5 year anniversary to complete the exit interview) and the average follow-up duration will be approximately four years. Participants who experience a stroke or MI during follow-up will be maintained on their randomized study drug and will continue to be followed for the occurrence of other primary and secondary outcomes until they are withdrawn at trial termination or are censored due to death.

To rigorously evaluate the effect of treatment on each defined secondary outcome, the strategy of retaining participants on treatment after a primary endpoint is considered as a requirement. A participant who experiences an MI at day 60 after randomization and a subsequent stroke at day 300 will be analyzed as an outcome for the primary analysis at day 60 (time to first stroke or MI). However, for the analysis of stroke alone, a participant will be counted as event free until day 300 (time to first stroke). This design allows for the detection of an effect of treatment

on stroke regardless of the treatment effect on the occurrence of MI. Similar contingencies pertain for the other secondary outcomes.

The occurrence of initial vascular events may affect a participant's ability and willingness to continue on his or her assigned study drug. Although all analyses will be conducted according to the intention-to-treat principle, an early effect of treatment on one type of outcome (e.g., stroke) may systematically alter compliance by treatment group, thus affecting the ability to detect an effect of treatment on later events. Because of this concern, every effort will be made to maintain each participant on his or her assigned treatment, both before and following any vascular event they may experience.

Interim Monitoring Plan

Interim monitoring is performed by the Statistical Center and focuses on recruitment (overall and by site), prevalence of insulin resistance, baseline comparability of treatment groups, protocol adherence, completeness of data retrieval, accrual of primary endpoint events (i.e., information accrual), safety and treatment efficacy. Recruitment, prevalence of insulin resistance, and completeness of data retrieval are also monitored by the Coordinating Center for purposes of daily trial operations and quality assurance. Independent and definitive monitoring by the Statistical Center; however, serves as a check on the Coordinating Center and provides the basis for reporting to the DSMB.

- *Monitoring for Recruitment and Prevalence of Insulin Resistance*

The Coordinating Center will monitor all steps in recruitment to assure early recognition of inadequate performance and identify reasons for inadequate performance in each recruitment site and in the trial overall. To assist in this process the Statistical Center will provide the Coordinating Center with data monitoring reports. These reports will include number of screening log entries, reasons for ineligibility on log, number of screener forms completed, reasons for ineligibility at screening, number of informed consent documents signed, number of screening HOMA tests obtained, the prevalence of insulin resistance (i.e., HOMA >3.0) among screened individuals, and randomizations. Each recruitment site's randomization performance is gauged in relation to its specific recruitment goal. The same reports are available at each DSMB meeting. Among the advantages of continuous monitoring of recruitment, the resulting data gives the DSMB an early and regular opportunity to compare the trial assumptions with the observed data to make early judgments about the merits of continuing the research. For example, within one year of initiation of funding, it is anticipated that 1506 will be screened with HOMA testing. The DSMB may consider stopping the trial or modifying the design if the observed prevalence of insulin resistance (HOMA >3.0) is substantially less than the assumed 45% and results in lower than expected rates of accrual and information (events).

- *Baseline Comparability*

  The adequacy of randomization is assessed by comparing the distribution of baseline characteristics among the treatment groups: demographics, physical and medical history measures, symptoms and complaints, prescription and non-prescription medications. Continuous variables are examined for skewness, outliers, or other departures from a normal distribution graphically and by summary statistics. F tests for equality of variances are used to determine the appropriate statistical test for continuous measures. Categorical variables are examined by calculating frequency distributions.

- *Protocol adherence*

  Protocol adherence is monitored monthly by the Operations Committee to assure early identification of poor performance at individual sites and in the trial overall. Periodic reports are provided to the DSMB and whenever requested. Specific parameters to be monitored include:

  - Randomization of ineligible participants
  - Treatment allocation errors (participant put on wrong treatment)
  - Failure to obtain required blood tests on time
  - Failure to complete required follow-up interviews on time
  - Loss and withdrawal rates
  - Treatment adherence (pill counts, cessation of use)
  - Failure to complete endpoint adjudication on time (within 30 days)
  - Use of TZDs outside of trial protocol
  - Continued use of trial medication despite contra-indication (e.g., hemoglobin <8.5 g/dl, ALT >2.5x upper limit of normal)

- *Monitoring for Efficacy*

  Three interim "looks" at the primary endpoint data for treatment efficacy are planned. The proposed stopping boundaries for the interim analyses are based on the work of Lan[1] and DeMets[2] and of Wang and Tsiatis[3] and are calculated using East Version 4. Since it is unlikely that there will be an adequate number of events during the first year to evaluate efficacy, interim "looks" are proposed after 227 events (one-fourth of total events), 454 events and 681 events are observed in the trial cohort, with the final analysis at 908 events, the total number of events selected for the trial. The stopping boundaries are designed to be flexible; therefore, the actual timing of the analyses will be based on Data and Safety Monitoring Board request. The Data and Safety Monitoring Board will review all interim analyses. If the interim analyses indicate a significant treatment effect or a very low conditional probability of detecting the estimated treatment effect (see Figure 1), the DSMB may choose to stop the trial prematurely.

- *Monitoring for Futility*

A futility analysis for the primary endpoint is planned during the 2nd and 3rd interim analyses (see Figure 1), which will occur after approximately 454 events of stroke or MI events have been observed and after 681 events. If either of these analyses shows that the nominal critical point (Z-score) for the primary endpoint crosses the internal boundary for futility, it would indicate that the observed effect size is much smaller than anticipated and that the trial has very low conditional power to detect the estimated treatment effect for the primary outcome. In addition to these futility analyses, the observed primary outcome event rate and rate of information accrual (number of events) is monitored from the initiation of the trial and compared with the expected rates. Based on these rates an estimated time for study completion (i.e. time to reach 908 events) and the estimated number of events that will be reached by the scheduled end of the trial is presented to the DSMB. When considering the futility analyses and rate of information accrual, the DSMB may also consider other internal evidence of futility (i.e., secondary outcomes) and has the option of recommending early termination of the trial or continuing with a possible adjustment to sample size.

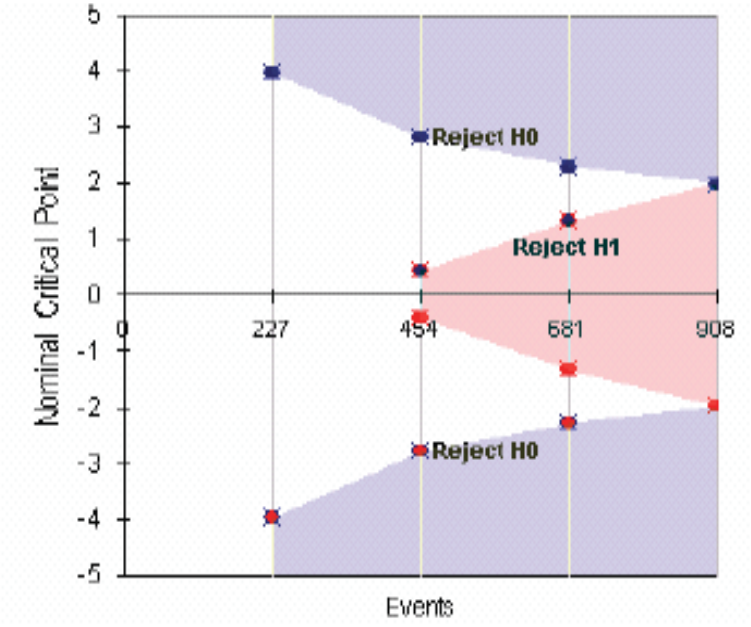

Figure 1. Proposed Stopping Boundaries for Three Interim Analyses. Interim analyses will be performed after 227, 454 and 681 endpoints are observed in the trial cohort. If the nominal critical point falls in the outer shaded are (indicating a significant treatment effect) or the inner shaded area (indicating a very low conditional probability of detecting the estimated treatment effect), the DSMB may choose to stop the trial prematurely.

- *Analysis of Adverse Events*

  The incidence of serious adverse events, as defined by the International Conference on Harmonization[4] and FDA guidelines, and the incidence of non-serious adverse events including headache, muscle aches, sinusitis, pharyngitis, teeth problems, edema, weight gain, shortness of breath, elevated ALT,) are calculated and compared among the treatment groups using statistics appropriate for discrete or count data. To track changes in weight over time, mean changes and the proportion of participants with an increase of 5% and 10% in body weight from baseline will be compared between the two treatment groups at each time point. In addition, the incidence of all cancers (and bladder cancer specifically), liver failure, serious congestive heart failure, serious macular edema, and bone fracture are analyzed by treatment group. Initially, due to the presence of small counts of serious and non-serious adverse events before data collection completion, the Fisher's exact test are used for comparing proportions with no (adverse) events with proportions with one or more (adverse) events among treatment groups. Finally, the effect of treatment on all-cause mortality as a time to event analysis is monitored as an additional safety endpoint.

Study Outcomes

The primary outcome measure is time to first fatal or non-fatal stroke or fatal or non-fatal MI. There are six secondary outcomes for this study: time to first stroke; time to first episode of Acute Coronary Syndrome (MI or unstable angina); all-cause mortality; the composite of time to first stroke, MI or serious CHF; time to first developing diabetes; and the modified mini mental state exam longitudinal scores. The date of randomization will be used as the time origin for a participant's study period. Time to censoring will be defined as time in days from date of randomization until the date of last contact. Contact will be defined as direct contact with the patient or a pre-defined surrogate. Database review will not be considered sufficient contact to determine event free time except in certain circumstances for all-cause mortality only (see below). Events detected through database searches will be counted if adjudicated as an event.

Participants are to be exited during an exit interview at the end of five years of study participation or at the last scheduled contact that falls between March 9, 2015 and July 7, 2015, whichever comes first.

Maximum follow-up will be 5 years plus a 21 day allowance for late completion of the exit interview (1847 days total).

Outcome and safety events will not be counted in the analysis if they occur after the date of completion of the exit interview. In addition, outcome and safety events will not be counted if they occur more than 21 days after the scheduled date of an exit interview that was not completed or completed late.

The IRIS trial will have endpoint review committees for stroke, acute coronary syndromes, and diabetes. Each committee will be comprised of a chair and three or more physicians who are specialists in the committee's content area. Reviewers will receive training in the application of specific criteria for declaring and classifying endpoints in their specialty. Reviewers will be blinded to treatment allocation. For suspected stroke endpoints, neurologists on the endpoint committee will decide whether a stroke has occurred and determine the stroke subtype. For suspected cardiac endpoints, cardiologists on the committee will decide if an acute coronary event has occurred and whether it was an MI or episode of unstable angina. The diagnosis of fatal MI includes participants who die after presentation with acute MI and cases of well documented sudden death. For suspected incident cases of diabetes, endocrinologists on the committee will decide if metabolic data fulfill the case definition.

The IRIS investigators will apply to the National Death Index (NDI) for information on patients who have been lost to follow-up, if approved by governing investigational review boards. At least two searches will be performed, one shortly after completion of enrollment and one approximately six months prior to the anticipated locking of the database. When an IRIS participant is found to have a matching record in the NDI, the record will be forwarded to the research team at the site which was managing the participant. That team will then perform a "follow-back" investigation by contacting the decedent's next of kin, physician, or hospital to confirm vital status and, if possible, the cause of death. All information on death or cause of death will be entered into the usual IRIS adjudication process. If the death is adjudicated or the cause of death is adjudicated as an MI or a stroke, the event will be included in the analyses for those outcomes. Any death identified from the NDI search on patients who were lost to follow-up that occurs after a participant's exit date, will be used as evidence of event free survival for all-cause mortality analysis only through the participant's exit date.
For complete protocol definitions of the primary and secondary outcomes see Appendix 1.

Analytical Plans

All final analyses will be conducted using SAS 9.3; R cmprsk will be used for the competing risk analyses.

- *Baseline Characteristics by Treatment*

  The adequacy of randomization is assessed by comparing the distribution of baseline characteristics among the treatment groups: demographics (age, gender, marital status, ethnicity, race, education); physical and medical history measures (weight, BMI, waist/hip ratio, abdominal obesity, systolic and diastolic blood pressure, edema, type of index event (stroke, TIA), presence of carotid angioplasty and carotid endarectomy before randomization, prior stroke before index event, MI, history of hypertension, history of CHF, atrial fibrillation,

coronary angioplasty, CABG, smoking status, history of coronary artery disease (CABG, angioplasty, or hospitalization for MI) modified MMSE score, NIH stroke scale score, modified Rankin grade; symptoms and complaints (headaches, muscle aches, sinusitis, pharyngitis, teeth problems, shortness of breath); laboratory values (HOMA, glucose, ALT, A1C, hemoglobin); prescription and non-prescription medications (aspirin, multivitamin, anti-hypertensives, loop diuretics, lipid lowering agents, anti-arrhythmic drugs, anti-coagulants, anti-platelets, other). Continuous variables will be examined for skewness, outliers, or other departures from a normal distribution graphically (side-by-side histograms, boxplots, normal quantile plots) and by side-by-side summary statistics (means, standard deviations, medians, quartiles, etc.). Inferences on differences in the means for continuous variables (e.g. age, BMI, weight) by treatment groups are accomplished through two-sample t tests. F tests for equality of variances are used to identify the appropriate statistical tests for continuous measures.

Categorical variables are examined by calculating frequency distributions. The chi-square test or Fisher's exact test (when cell counts are less than 5) are used for comparing the distribution between treatment groups. Except for medical history questions, analyses are run with "Unknown", "Uncertain", and "Refused" responses coded as missing and "Missing" excluded from the analysis. For the following medical history variables, the analyses are run with "Unknown" and "Refused" coded as missing (excluded from the analysis) and 'Uncertain" responses counted as 'no' or 'absent' responses: Prior stroke before index event, MI, history of hypertension, history of CHF, coronary angioplasty, CABG, smoking status, symptoms and complaints (headaches, muscle aches, sinusitis, pharyngitis, teeth problems, shortness of breath).

- *Analysis of Primary Outcome*

All analyses will be by intention-to-treat. That is, participants will be analyzed according to their original treatment assignment, regardless of adherence. The analysis will not be adjusted for site because of the large number of sites and the few numbers of participants enrolled at many of the sites. The analysis of the primary endpoint, time to first stroke or MI, will be tested by the log rank statistic[5], with an experiment-wide type I error of 0.05 (2-sided) as the level of significance. Cumulative event-free rates for the primary endpoint, as well as the individual components (stroke and MI), will be calculated using the method of Kaplan-Meier. Cox proportional hazards models will be used to examine the effect of treatment[6]. The proportional hazards assumption will be tested by treatment by time interaction term. Breslow's approximation, which is the default in PROC PHREG in SAS, will be used to handle tied event times when ties are relatively few. If data are heavily tied we will use Efron's method, which gives closer results to the exact results than Breslow's approximation. The treatment effect will be summarized as a hazard ratio (pioglitazone vs. placebo) with 95% confidence interval adjusted for the interim looks.

Because of the size of this study, we expect that the randomization process will produce comparable groups of participants and balance. However, in a secondary sensitivity analysis, we will adjust for the following pre-specified set of baseline variables that may affect prognosis or treatment effect: age, gender, systolic blood pressure, diastolic blood pressure, history of hypertension, type of entry event (i.e., TIA or ischemic stroke), history of prior stroke before the index event, history of coronary artery disease (history of CABG, angioplasty, or hospitalization for MI), and current smoking. In addition, secondary analyses of the effect of treatment in predefined baseline subgroups will be examined by tests of interaction adjusted for multiplicity. The pre-specified subgroups are:

- Age in years
- Gender
- Race (defined as white, black and other)
- Ethnicity (Hispanic vs. non-Hispanic)
- HDL cholesterol
- Fasting plasma glucose
- HOMA
- HbA1c
- Triglyceride
- BMI
- History of hypertension
- History of coronary artery disease (CABG, Angioplasty, or hospitalization for MI)
- Medication adherence (percent of mgs taken based on protocol dosage)
- Medication adherence (<80% vs. ≥80% of mgs taken based on protocol dosage).

Cox proportional hazards models will be used to examine the effect of treatment in these adjusted and subgroup analyses[6]. The Hochberg procedure, a sequentially rejective variation of the Bonferroni procedure, will be used to determine significance using an overall type I error of 0.05 (two-sided)[7]. Subgroups analyzed as a continuous measure that are determined to significantly interact with treatment will be further explored based on established clinical cutpoints.

Additional subgroup analyses that will be considered for exploratory tertiary analyses are baseline CRP, baseline statin use, change in HOMA from baseline to year 1, change in CRP from baseline to year 1, change in HbA1c, baseline triglyceride/HDL ratio, baseline non-HDL cholesterol (total cholesterol-HDL cholesterol), atrial fibrillation, and subtype and territory of index event.

- *Analysis of Secondary Outcomes: Stroke, ACS, All-Cause Mortality, Stroke, MI or serious CHF, Diabetes*

Any stroke, acute coronary syndrome, all-cause mortality, stroke, MI or serious CHF, and progression to diabetes will be analyzed as time to first event in the same manner as the primary endpoint principal analysis. The treatment effect will be summarized as a hazard ratio (pioglitazone vs. placebo) with 95% confidence interval. To provide some control for multiplicity of secondary outcomes, the Hochberg procedure will be used to determine significance using an overall type I error of 0.05 (two-sided)[7].

- *Sensitivity Analyses*

In addition, depending on the availability of resources, several sensitivity analyses of IRIS outcomes may be conducted using updated criteria that have evolved since initiation of the study (see Appendix 1):

(1) Analysis of the effect of treatment on risk of the primary outcome (stroke or MI) using adjudicated MI outcomes based on the 2012 ESC/ACCF/AHA/WHF criteria for MI[8] and the 2013 AHA/ASA criteria for stroke[9].
(2) Analysis of the effect of treatment on risk of stroke using adjudicated stroke outcomes based on the 2013 AHA/ASA criteria.
(3) Analysis of the effect of treatment on risk of ACS using adjudicated MI outcomes based on the 2012 ESC/ACCF/AHA/WHF criteria for MI.
(4) Analysis of the effect on treatment on risk of the composite safety outcome (stroke, MI, serious CHF) with new MI based on the 2012 ESC/ACCF/AHA/WHF criteria and new stroke criteria based on the 2013 AHA/ASA criteria.
(5) Analysis of the effect of treatment on progression to diabetes using adjudicated diabetes outcomes based on the revised 2010 ADA criteria for diabetes[10].

- Competing Risks Sensitivity Analysis

The occurrence of individual secondary outcomes (stroke, ACS, stroke, MI or serious CHF, and diabetes) may be circumvented by death. Naively censoring patients who fail from all-cause mortality in analyzing time to first occurrence of any of these secondary outcomes could lead to biased estimates. The Cox model commonly used in classical survival analysis is a valid analysis for cause-specific hazard in the presence of competing risks. But if we are interested in the cause-specific cumulative incidence, the Gray[11] and Fine & Gray[12] methods are more appropriate.

A competing risks sensitivity analysis will be conducted using these two methods. More specifically, Gray's test for subdistribution hazards will be used for comparing the cumulative incidences of each secondary outcome by treatment group[11] and the proportional subdistribution hazards model proposed by Fine & Gray[12] will be used to estimate the treatment effects (subdistribution hazard

ratios). The methodologies described above are available in and will be implemented through the R *cmprsk* package.

Gray's test will be compared with the logrank test for the treatment effect. We will also compare the subdistribution hazard from Fine-Gray model to the cause specific hazard from the Cox model to improve understanding of the treatment effect. A cumulative incidence curve will be compared to Kaplan-Meier curve.

- *Analysis of Secondary Outcomes: Cognitive Decline (Modified MMSE Score)*

  To determine whether pioglitazone, compared to placebo, reduces the risk for cognitive decline, as measured by the modified MMSE score, a longitudinal analysis will be done. Repeated measures of the modified mini mental state exam will be analyzed by mixed model methods adjusted for the study design[13]. In addition, secondary analyses will be conducted comparing participants who decline from a normal modified MMSE score (≥90) to an exit score of <78 by treatment groups using a logistic model.

  The primary analysis will be based on an analysis of covariance mixed model approach using maximum likelihood estimation assuming missing data are missing at random. All participants with available data will be included in the model. Initially, we will determine the most appropriate covariance structure for the repeated outcomes and then fit the mean model. Treatment by time interactions will be examined to determine whether the treatment effect is constant over time. If there is evidence of an interaction ($p<0.05$), the data will be analyzed by time point.

  - Sensitivity Analysis for Cognitive Data

    The mixed effects model based on maximum likelihood estimation is unbiased if missing data are unrelated to outcomes, i.e., if the data are considered missing at random (MAR) or missing completely at random (MCAR). Nevertheless, because it is not always known whether missing data are ignorable and because missing observations have the potential to alter the results of analyses, the pattern of missing data and dropouts will be examined among the two treatment arms. We will also examine whether the outcome is related to missingness by using logistic regression models to determine if the outcome measure at the follow-up time preceding the missed visit predicts the missing value and if baseline value predicts non-intermittent missing (i.e., missing last expected score). If there are no systematic differences between those with and without missing data, the data can be considered to be observed (and missed) at random. If the outcome measure is predictive of missing data, it would indicate that the missing data are potentially not MAR or MCAR. Also, if there are no covariates predictive of missing data it would be another indication that the data may likely be MAR. If the missing data are determined to be possibly non-ignorable and there is a reasonable likelihood that the conclusions could be altered by the missing information, sensitivity

analyses will be considered to determine the impact of missing information on the treatment comparisons including the approach proposed by Carpenter et al.[14] (e.g., multiple imputation that reflects the behavior of the missing data or including variables that are predictive of missingness in the model).

SAS code for all completed analyses is available upon request.

References

1. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659-63.
2. DeMets DL, Lan KKG. Interim analysis: the alpha spending function approach. Statistics in Medicine 1994;13:1341-52.
3. Wang SK, Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. Biometrics 1987;43:193-99.
4. Dixon JR. The International Conference on Harmonization Good Clinical Practice Guideline. Quality Assurance 1998;6:65-74.
5. Kalbfleish J, Prentice R. The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons; 1980.
6. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society 1972;34:187-202.
7. Hochberg Y. A sharper Bonferroni Procedure for multiple tests of significance. Biometrika 1988;75:800-2.
8. Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. Journal of the American Academy College of Cardiology. 2012;60:2020-35.
9. Sacco RL, Kasner SE, Broderick J, et al. An updated definition of stroke for the 21st century. A statement for healthcare professionals from the American Heart Association/American Stroke Association Stroke Council. Stroke 2013;44:2064-89.
10. Association AD. Standards of medical care in diabetes - 2010. Diabetes Care. 2010;33:S11-S61.
11. Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. The Annals of Statistics, 1988; 16(3): 1141-1154.
12. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, 1999; 94(446):496-509.
13. Diggle PJ, Liang K-Y, Zeger SL. Analysis of Longitudinal Data. Oxford: Clarendon Press; 1994.
14. Carpenter J, Pocock S, Lamm CJ. Coping with missing data in clinical trials: A model-based approach applied to asthma trials. Statistics in Medicine, 2002; 21:1043-1066.

.

Appendix 1: Endpoint Definitions

A. Protocol Endpoint Definitions (Primary and Secondary Outcomes)

Fatal or non-fatal stroke or fatal or non-fatal MI were selected as the primary outcomes for the IRIS trial for three main reasons. First, they are the major vascular events contributing to morbidity and mortality after an initial ischemic stroke or TIA. Second, they have both been associated with insulin resistance, and treating insulin resistance may reduce risk for each of them. Third, stroke and MI can be detected and verified with a high degree of accuracy. In addition, we will examine several endpoints other than stroke and MI as secondary endpoints. Definitions for all primary and secondary endpoints are as follows:

Stroke

For purposes of this trial, stroke is an acute neurological event with focal signs or symptoms lasting at least 24 hours which represent a focal loss of brain function that can be attributed to a disturbance in one vascular distribution and for which no other cause is found. In addition, there must be at least a one-point increase in the NIH stroke scale in a previously normal section, or an appropriate new or extended abnormality seen on CT or MRI. IRIS will count non-traumatic intracerebral and subarachnoid hemorrhage as outcomes in addition to ischemic events. Subdural and epidural hematoma are not included as part of the primary stroke outcome. A stroke is classified as fatal if death occurs within 30 days or if, in the opinion of the neurology review committee, death is a direct result of a physiological consequence of a stroke event.

Myocardial Infarction

Acute myocardial infarction will be diagnosed according to criteria modified from the 2000 Consensus Conference of the European and American Colleges of Cardiology[1]. MI events will be diagnosed based on patient symptoms and electrocardiogram changes, in conjunction with contemporary biochemical markers of myocardial necrosis (troponin or CK). Both acute ST segment elevation and non-ST segment elevation MI will be combined as an outcome. An acute, evolving, or recent non-fatal MI will be diagnosed in the presence of a typical rise and gradual fall (troponin) or more rapid rise and fall (CK-MB) of biochemical markers of myocardial necrosis with at least one of the following: a) ischemic symptoms, b) development of pathologic Q waves on the ECG, c) ECG changes indicative of ischemia (ST segment elevation or depression), d) coronary artery intervention (e.g., angioplasty), e) angiographic findings indicative of unstable plaque or thrombus in a coronary artery, or f) cardiac imaging (perfusion or echocardiogram) suggestive of MI. Criteria for biochemical markers depend on whether the event was (a) spontaneous, or in setting of (b) cardiac catheterization, angioplasty or stenting, or (c) CABG. Biochemical markers will be considered indicative of necrosis when the CK-MB exceeds (a) 1.5, (b) 3.0 and (c) 5.0 x ULN, respectively, or troponin concentration exceeds (a) 2.0, (b) 5.0, and (c) 10 x ULN, respectively. For participants who do not seek medical care at the time of acute symptoms, a recent non-fatal MI will be diagnosed in the presence of a history of ischemic symptoms and new Q waves on ECG or imaging data clearly indicative of new, focal myocardial damage.

A fatality will be classified as "fatal MI" according to criteria from the Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/TexCAPS),[2] which requires that there be no non-cardiac cause of death and one of the following: death within 30 days from the onset of symptoms of a definite acute myocardial infarction; witnessed unexpected sudden cardiac death within 1 hour of symptoms; death occurring > 1 hour but < 24 hours after collapse; and unwitnessed and unexpected death that is presumed to be sudden (in this case, there must be confirming autopsy data or, preceding history of CHD events or symptoms).

Acute Coronary Syndrome/Unstable Angina

Acute coronary syndrome (unstable angina, fatal or non-fatal myocardial infarction) are a secondary outcome. The diagnosis of unstable angina will be based on the well-defined criteria from the AFCAPS/TexCAPS study[2]. This diagnosis requires the absence of a diagnosis of myocardial infarction and the presence of new-onset, accelerated exertional or rest chest pain (anginal in quality) lasting > 5 minutes and at least one of the following: mildly elevated troponin values (<2 x upper limit of normal), new or presumably new ST segment deviation >0.5mm in 2 or more leads, T-wave inversion >3mm in 3 leads, left bundle branch block, or abnormal stress testing. Abnormal stress testing is defined as >1mm horizontal or down-sloping ST-segment depression in two or more leads with evidence of abnormalities in stress imaging or coronary angiography as described below (if imaging or angiography is obtained).

Stress imaging evidence:

- Perfusion imaging (MIBI or thallium): A reversible perfusion defect with exercise or pharmacologic stress (adenosine, dipyridamole, dobutamine) imaging; or
- Echocardiography: An inducible regional wall motion abnormality after exercise or pharmacologic stress.

Angiographic evidence:

- >70% stenosis in the LAD, circumflex, or RCA or one of their major branches, or > 50% left main coronary artery stenosis.

Diabetes Mellitus

Diabetes mellitus will be classified based on 2004 American Diabetes Association guidelines[3]. According to these guidelines diabetes is diagnosed when one of three conditions is met: 1) the value of two fasting glucose measurements equal or exceed 126 mg/dl, 2) two random plasma glucose equals or exceeds 200 mg/dl in the presence of typical symptoms of diabetes, or 3) the plasma glucose equals or exceeds 200 mg/dl two hours after a 75 gram oral glucose load, confirmed on a repeat test. In practice, the oral glucose tolerance test is rarely used. For this reason and to reduce the burden of participating in the trial, we will classify a participant as having diabetes if one of the first two criteria is met. In addition, IRIS participants will be classified as diabetic in the presence of compelling indicators of hyperglycemia, even if requisite ADA testing is not performed. Specifically, patients will be classified as diabetic if any of the following three conditions is met:

- A physician diagnoses diabetes and prescribes an antidiabetic agent and one of the following test results is documented:
  - Fasting plasma glucose concentration equals or exceeds 126 mg/dl (7.0 mmol/L);
  - Random plasma glucose concentration equals or exceeds 200 mg/dl (11.1 mmol/L);
  - 2-hour OGTT glucose equals or exceeds 200 mg/dL (11.1 mmol/L);
  - HbA1c equals or exceeds 7.0%.
- Two fasting glucometer (capillary blood) glucose values obtained in a health-care setting (i.e., not a home test) equal or exceed 151 mg/dL (8.4 mmol/L) (i.e., 1.2 x 126 mg/dL [7 mmol/L]);
- Two random glucometer (capillary blood) glucose values obtained in a health-care setting (i.e., not a home test) equal or exceed 240 mg/dL (13.2 mmol/L) (i.e., 1.2 x 200 mg/dL [11 mmol/L]) in the presence of typical symptoms of hyperglycemia (e.g., polyuria, polydipsia, weight loss, blurry vision); In the situation where blood glucose levels are only available from a glucometer (capillary blood), it is acceptable to use these data, but only if the test is performed in a healthcare setting (physician's office, clinic, or skilled nursing facility), since proper standardization and use of the meter cannot be assumed with at-home testing. To be conservative, a margin of error of +20% has been incorporated, since the minimum acceptable accuracy for glucometer results is within 20% at glucose concentrations over 75 mg/dL.

A diagnosis of diabetes that is made during a hospital admission will not be recognized unless one of the following two conditions is met:

1. The hospital admission was for a diabetic emergency, such as diabetic ketoacidosis or hyperosmolar hyperglycemic syndrome;
2. The patient was discharged on any anti-hyperglycemic agent and a HbA1c ≥7.0% was obtained during the hospitalization.

All-Cause Mortality

All-cause mortality will serve as a routine check on safety. It is also a very objective, additional measure of treatment effectiveness.

Cognitive Decline

Decline in cognitive functioning will be measured as a change in mean score in the Modified Mini Mental State Examination (3MS). In a secondary analysis, we will restrict the analysis to participants who entered IRIS with a normal 3MS (>90) and define a significant decline as an exit score below 78, the customary criterion for significant impairment318. The reason to restrict this secondary analysis to participants with a normal baseline score is that any discrete change in score may have different importance for people with different baseline function. A fall of 12 points from normal to less than 78, however, is generally regarded as significant.

Severe Congestive Heart Failure

CHF episodes that result in hospital admission or death will be counted in a combined secondary outcome (along with stroke and MI). The composite outcome was added at the request of the Data Safety and Monitoring Board (DSMB) to create a measurement of the net benefit of pioglitazone. Episodes of CHF are classified by members of the Cardiology Outcome Committee. In completing its assessment, the Committee will be guided by usual practices in diagnosing heart failure that require consideration of typical symptoms (shortness of breath, weight gain, edema, fatigue, dyspnea on exertion, PND), signs (weight gain, pulmonary râles, JVP, edema, tachypnea, tachycardia), test findings (e.g. pulmonary edema on chest radiography, serum BNP concentration) and alternative diagnoses. Non-fatal episodes of CHF leading to emergency room visits or hospital stays less than 24 hours in duration will not be counted as secondary outcomes. (All adjudicated episodes of CHF, regardless of severity or resultant hospitalization, are reported to the DSMB with other adverse events.)

B. Endpoints for Sensitivity Analyses Definitions

Stroke: 2013 AHA/ASA Criteria for Stroke

At the time the IRIS trial was initiated in 2004, it used a definition for stroke that was widely accepted at the time and based on clinical and imaging data (see section A above). The definition of stroke includes non-traumatic subarachnoid hemorrhage and intracerebral hemorrhage. To account for advances in science and technology, in 2013, the American Heart Association/American Stroke Association issued a Statement for Healthcare Professionals describing an updated definition for stroke[5]. The update was motivated by opinion that the older, standard WHO definition had been made obsolete by evidence that permanent injury can result from events lasting less than 24 hours. In addition, the AHA/ASA felt that a uniform definition was needed for purposes of clinical care and research. The resulting new AHA/ASA definition for stroke broadly includes symptomatic and asymptomatic events caused by CNS infarction, intracerebral hemorrhage, subarachnoid hemorrhage, and cerebral venous thrombosis. There is also a category of "stroke, not otherwise specified" that is defined by an episode of acute neurological dysfunction presumed to be caused by ischemia or hemorrhage, persisting ≥ 24 hours or until death, but without sufficient evidence to be classified as one of the above." The most significant feature of the AHA/ASA updated definition, and the only feature that affects IRIS, relates to the definition of ischemic stroke as "an episode of neurological dysfunction caused by focal cerebral, spinal, or retinal infarction." The definition for CNS infarction is based on:

1. pathological, imaging, or other objective evidence of cerebral, spinal cord or retinal focal ischemic injury in a defined vascular distribution; or
2. clinical evidence of cerebral, spinal cord, or retinal focal ischemic injury based on symptoms persisting ≥24 hours or until death, and other etiologies excluded (Note: CNS infarction includes hemorrhagic infarctions, types I and II).

Imaging evidence for infarction, according to the updated AHA/ASA criteria, includes findings from MRI or CT. The statement recognizes that MRI is more sensitive than CT scanning, particularly within first 12 hours, but that neither is perfectly accurate. This

updated definition, if applied to participants in the IRIS trial, would likely result in more events being classified as stroke compared with the original IRIS definition. This is because the original IRIS definition requires symptoms or signs to last at least 24 hours, even in the presence of imaging evidence for infarction. The updated definition classifies an episode of focal neurological dysfunction as an ischemic stroke if there is evidence for infarction regardless of duration of symptoms.

The updated definition for stroke caused by ICH ("Rapidly developing clinical signs of neurological dysfunction attributable to a focal collection of blood within the brain parenchyma or ventricular system that is not caused by trauma") is identical to the IRIS definition, including the exclusion of subdural and epidural hematomas. The updated definition for stroke caused by SAH ("Rapidly developing signs of neurological dysfunction and/or headache because of bleeding into the subarachnoid space {the space between the arachnoid membrane and the pia mater of the brain or spinal cord}, which is not caused by trauma.") is also identical to the current IRIS definition.

The ancillary analysis for the primary outcome will define fatal and non-fatal stroke according to the AHA/ASA updated definition for stroke and represents a sensitivity analysis to show readers how our findings are affected by use of the more modern definition. It is important to note that the primary outcome for the IRIS trial (i.e., fatal or non-fatal MI plus fatal or non-fatal stroke) does not change and will still be classified and reported using the original IRIS protocol and its older stroke definition. However, the sensitivity analysis will be reported to show the effect of the new criteria on IRIS findings.

Myocardial Infarction: 2012 ESC/ACCF/AHA/WHF criteria for MI

Since the IRIS trial was designed in 2004, two international consensus panels have revised the definition of myocardial infarction (MI). In 2000, the First Global MI Task Force had proposed that any troponin elevation in the setting of myocardial ischemia be interpreted as an indicator of necrosis and MI, but there was little early acceptance of this definition[1]. In 2007, the Joint ESC/ACCF/AHA/WHF Task Force reiterated this definition in the Universal Definition of Myocardial Infarction Consensus Document and further distinguished between different conditions that lead to MI (e.g., type 1=spontaneous, type 2=imbalanced in balance of myocardial oxygen demand and supply, type 3=sudden unexpected cardiac death, type 4=infarction related to PCI or stent thrombosis, and type 5=infarction related to CABG)[6]. For spontaneous MI, the panel based the diagnosis on a rise or fall of cardiac biomarkers (preferably troponin) with at least one value above the 99th percentile upper reference limit (URL). This new cutoff for cardiac biomarkers was lower than the cutoff of 1.5 to 2.0 x 99th percentile URL that had been commonly used in the past and was used for the IRIS primary outcome definition. The revised MI definition recognized progress in standardizing and improving cardiac troponin (cTn) as a specific biomarker of myocardial necrosis. The second Joint ESC/ACCF/AHA/WHF panel published the Third Universal Definition of Myocardial Infarction in 2012. This document further refined the definition of MI with the following additional changes:

- Addition of "identification of an intracoronary thrombus by angiography or autopsy" in the list of evidence that could support the diagnosis of acute MI.
    - This criterion is consistent with the current IRIS criteria.
- Minor changes in the criteria for qualifying ST segment elevation based on gender and age. The revised criteria require new ST elevation at the J point in two contiguous leads with the cut-points: >0.1 mV in all leads other than leads V2 or V3 where the following cut points apply: >0.2 mV in men >40 years; >0.25 mV in men <40 years, or >0.15 mV in women.
    - These criteria have slight differences from the current IRIS criteria.
- Revision of criteria for the diagnosis for sudden cardiac death or cardiac death when biomarkers are unavailable to require pre-morbid symptoms suggestive of myocardial ischemia and to remove mention of evidence from angiography or autopsy. The new criteria represent a gradual movement to require more specific evidence for a cardiac etiology before attributing sudden or unexpected death to myocardial infarction.
    - These criteria are more conservative than the current IRIS criteria, which attribute death to a MI if it is unwitnessed, there is no other identified cause, and it is presumed to be sudden (in this case there must be confirming autopsy data or, if an autopsy was not performed, a preceding history of CHD events or symptoms).
- Change in the criterion for the diagnosis of MI in the setting of percutaneous coronary intervention from an elevation of cardiac biomarker values (preferably cTn) >3x 99th percentile URL to >5 x 99th percentile URL in patients with normal baseline values or a rise of cardiac biomarker values >20% if the baseline values are elevated and are stable or falling. In addition, the diagnosis requires one of several other signs of myocardial injury: symptoms of ischemia, new ischemic ECG changes, persistent loss of patency of a coronary artery or side-branch, or imaging evidence of loss of function or viability.
    - These cTn criteria (>5 x 99th percentile URL) are the same as for IRIS. However, the IRIS criteria also allow for CK-MB >3x 99th percentile. IRIS does not require any other symptom or sign of myocardial injury in this setting.
- Change in the criterion for the diagnosis of MI in the setting of coronary bypass surgery from an elevation of cardiac biomarker values >5 x 99th percentile URL to > 10 x 99th percentile URL. In addition, the diagnosis requires one of several other signs of myocardial injury: pathologic Q wave or LBBB, new graft or native vessel occlusion, or imaging evidence of loss of function or viability.
    - These new cTn criteria (> 10 x 99th percentile URL) are the same as the current IRIS criteria. However, the IRIS criteria also allow CK-MB >5 x 99th percentile URL when cTn is not available.
    - IRIS does not require any other signs or symptoms of myocardial injury in these settings.

A summary of similarities and differences between the Third International Definition and the IRIS definition appear in Table I.

The 2012 revisions are expected to become the standard for diagnosis of myocardial infarction, although controversy continues about the clinical importance of infarcts associated with minimal troponin release. Troponin is an extremely sensitive marker of myocardial injury and minimal levels of troponin release do not lead to impairment in cardiac function. Nonetheless, most studies are adopting the new criteria. To ensure comparability of the IRIS results to other research, we will perform an ancillary analysis to determine the impact of the Third International Definition of MI on IRIS findings.

The ancillary analysis for the primary outcome will define fatal and non-fatal MI outcomes according to the newly revised definition for MI and represents a sensitivity analysis to show readers how our findings are affected by use of the more modern definition. It is important to note that the primary outcome for the IRIS trial (i.e., fatal or non-fatal MI plus fatal or non-fatal stroke) does not change and will still be classified and reported using the original IRIS protocol and its older MI definition. The sensitivity analysis will be reported to show the effect of the new criteria on IRIS findings.

Table 1. Comparison of Third International Definition for MI and IRIS Criteria for MI		
Item	3Rd International	IRIS
Enzyme threshold for Dx of Spontaneous MI	> 1x 99th upper percentile URL	Troponin > 2x URL or CB-MK > 1.5 x URL
Symptoms	Symptoms of myocardial ischemia	Symptoms of myocardial ischemia
ECG: Q-Waves	New Q waves ≥ 0.03 sec and ≥0.1 mV deep or QS complex in leads I, II, aVL, aVF or V_4-V_6 in any two leads of a contiguous lead grouping (I, aVL; V_1-V_6; II, III, aVF).	New Q waves >0.03 sec in 2 or more contiguous leads.
ECG: LBBB	New LBBB	New LBBB
ECG: ST elevation	New ST elevation at the J point in two contiguous leads with the cut-points: ≥0.1 mV in all leads other than leads V_2 – V_3 where the following cut points apply: ≥0.2 mV in men ≥40 years; ≥0.25 mV in men <40 years, or ≥0.15 mV in women.	New ST elevation at the J point in two or more contiguous leads with the cut-points ≥0.2 mV in leads V_1 – V_3 and ≥0.1 mV in all other leads.
ECT: ST depression	New horizontal or down-sloping ST depression ≥0.05 mV in two contiguous leads and/or T inversion ≥0.1 mV in two contiguous leads with prominent R wave or R/S ratio >1.	New/presumed new ST depression ≥0.1 mV and/or T inversion ≥0.1 mV in two or more contiguous leads
IC Thrombus	Angiographic findings or autopsy indicative of thrombus in coronary artery	Angiographic findings indicative of thrombus in a coronary artery or unstable plaque.
Imaging	Imaging evidence of new loss of viable myocardium or new regional wall motion abnormality	Cardiac imaging clearly indicative of new focal myocardial damage (fixed perfusion defect or echo wall motion abnormality)
Autopsy	Autopsy findings of coronary thrombus.	Autopsy evidence of an acute MI.

Diabetes: 2010 ADA Criteria for Diabetes

In 2010, the ADA revised the criteria for diabetes to include hemoglobin A1c > 6.5% (confirmed by repeat testing). To determine how the IRIS results would be affected by use of this revised definition, we will conduct a sensitivity analysis using the revised criteria. No other aspects of the IRIS DM outcome classification are affected.

Appendix References:

1. The Joint European Society of Cardiology/American College of Cardiology Committee. Myocardial infarction redefined - a consensus document of the joint European Society of Cardiology/American College of Cardiology Committee for the Redifinition of Myocardial Infarction. J Am Coll Cardiol 2000;36:959-69.
2. Downs JR, Beere PA, Whitney E, et al. Design and rationale of the Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/TexCaps). Am J Cardiol 1997;80:287-93.
3. Association AD. Diagnosis and classification of diabetes mellitus. Diabetes Care 2004;27 (supplement 1):S5-S10.
4. Bland RC, Newman SC. Mild dementia or cognitive impairment: The modified Mini-Mental State Examination (3MS) as a screen for dementia. Can J Psychiatry 2001;46:506-10.
5. Sacco RL, Kasner SE, Broderick J, et al. An updated definition of stroke for the 21st century. A statement for healthcare professionals from the American Heart Association/American Stroke Association Stroke Council. Stroke 2013;44:2064-89.
6. Thygesen K, Alpert JS, White HD, On behalf of the Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. Journal of the American Academy College of Cardiology 2007;50:2173-95.