ClinicalTrials.gov Identifier: NCT01215253

Late Sodium Current Blockade in High Risk ICD Patients (RAID)

STATISTICAL ANALYSIS PLAN

Authors:

Chris Beck, PhD, University of Rochester

Jack Hall, PhD, University of Rochester

January 4, 2017

Version 2.0

Revision History

Date	Version	Description
May 14, 2014	1.0	Original release of the document. Included the statistical analysis plan specified in the protocol, and added a plan for handling missing data in primary analysis (section 3.1, paragraph 2).
January 4, 2017	2.0	Added a plan for maintaining the significance level at 5% in the primary analysis should the trial terminate without hitting a stopping boundary (section 3.1, paragraph 3).

Table of Contents

1 Study Objectives

1.1 Primary Objective

The primary objective of the study is to determine whether ranolazine administration will decrease the likelihood of a composite arrhythmia endpoint consisting of ventricular tachycardia or ventricular fibrillation (VT/VF) requiring antitachycardia pacing (ATP), ICD shocks, or death.

1.2 Secondary Objectives

The secondary objectives of this trial are as follows:

1. to determine whether ranolazine administration will decrease the likelihood of a composite arrhythmia endpoint consisting of VT or VF requiring ICD shock or death (while excluding VT/VF requiring just ATP);
2. to determine whether ranolazine administration will decrease the likelihood of composite primary endpoints consisting of hospitalization for cardiac causes (including not only hospitalization for heart failure, but also hospitalizations related to cardiac arrhythmias, myocardial infarction or ischemia) or death, whichever occurs first;
3. to determine whether ranolazine administration will decrease the likelihood of a composite secondary endpoint consisting of CHF hospitalization or death;
4. to determine whether ranolazine therapy will decrease the number of repeated hospitalizations for cardiac causes;
5. to determine whether ranolazine administration will decrease the likelihood of repeated ICD therapies (not just first therapy);
6. to determine whether ranolazine administration will decrease the likelihood of inappropriate shocks (a decrease in episodes of atrial fibrillation triggering inappropriate therapy) evaluating the risk of first and risk of repeated inappropriate shocks;
7. to determine whether ranolazine therapy will be associated with improvement in exercise capacity measured by the 6-minute walk test (6MWT) and in the quality of life measured by the Kansas City Cardiomyopathy Questionnaire (KCCQ);
8. to evaluate the safety of ranolazine therapy utilizing ICD interrogation data documenting all types of ventricular tachyarrhythmias (including torsade de pointes).

2 Study Design

2.1 Endpoint Event Rates

Based on recent preliminary unpublished data from the MADIT Risk Stratification Study and from the MADIT-CRT trial, we revised our estimate of the 2-year cumulative endpoint rate to be 25% in the placebo arm of the trial (instead of original 30%). We maintain the same level of reduction of primary events: for the ranolazine arm (R), we expect at least a 25% reduction in risk, after allowing for the possibility of 10% cumulative crossovers to placebo (within 2 years) – that is, a hazard ratio (HR) of 0.75. This implies a 2-year cumulative event rate in R of 19% [= $1 - (1 - 0.25)^{0.75}$], representing a 24% reduction in 2-year cumulative event rates. We expect no crossovers from P to R. We allow for losses to follow-up at the rate of 5% per year.

2.2 Significance Level and Power

The null hypothesis is that the true cumulative probability curves for time to first endpoint event are identical in the R and P arms (implying a HR of 1.0). The alternative hypothesis is that the curve for the R arm is below that for the P arm, implying a reduction in risk of a first endpoint event. Power computation is focused on a constant hazard ratio (HR) of 0.75, a 25% reduction in ongoing risk, an amount deemed worthy as clinically relevant, although a HR of 0.667 is considered quite possible. The trial is designed to have a significance level of 0.05 (2-sided) and 80% power at a constant HR of 0.75. See the table below for resulting power at other HRs; power is 98% at a HR of 0.667.

2.3 Recruitment and Randomization

We will randomize 1440 patients in approximately 100 centers. The randomization is expected to require 33 months at an average rate of 44 per month, or 0.54 patients per center per month. Consented patients will be randomized equally to the R and P arms, with randomization stratified by enrolling center, by type of device (ICD vs. CRT-D), within center, and by prior history of VT/VF/cardiac arrest within device type. No further stratification is really feasible. We assume balance between treatments within each device type is more critical than within ischemic and non-ischemic patient groups, or other possible risk categories.

2.4 Sequential stopping rule

A trial that continues follow-up until a pre-specified number of endpoints have accumulated, and based on a log rank test, would require 380 endpoint events. At the anticipated randomization rate, loss rate and endpoint event rates, a larger sample size and/or longer follow-up time would be required.

We have chosen instead a sequential stopping rule with triangular stopping boundaries, similar to that used in the MADIT and MADIT-II trials.[1] The expected number of events needed to reach a termination boundary is greatly reduced, although some risk of a longer trial is encountered. We chose a specific boundary as provided by PEST software[2] (as well as our own) that would meet the significance level and power requirements.

The trial is to be monitored by periodically fitting a proportional-hazards regression model with treatment arm and six presumed risk factors as covariates (see Section 3.1 below) and stratified by enrolling center. The resulting score-test statistic, Z, for testing nullity of a treatment effect, when plotted against its variance V, behaves like a Brownian motion with β, the regression coefficient for treatment effect, as its drift [3] The statistic Z quantifies any difference between treatment arms R and P in estimated time-to-endpoint curves. Its variance V (roughly equal to the accumulated event count divided by 4) quantifies statistical information. This computation will be done monthly, once 20 events have been accumulated, and submitted to the DSMB chair and statistician. A plot of Z versus V, starting at the origin, will continue until it reaches one of the two boundaries:

$$14.8153 + 0.101103*V \text{ (upper)} \quad \text{and} \quad -14.8153 + 0.303309*V \text{ (lower)}$$

and truncated at $V = 130$. (The boundaries may be adjusted to recognize the discrete-time monthly monitoring.[3]) Upon reaching a boundary, the trial is to be terminated in favor of R if Z reaches the upper boundary and therefore $p \le 0.05$. If Z first reaches the lower boundary with $p \le 0.05$ (that is, prior to $V = 18.585$), the trial will be terminated with a conclusion that P is superior to R – i.e., that ranolazine increases risk of endpointing. If Z reaches the lower boundary with $p > 0.05$ ($V > 24.24$), or along the vertical strip at $V = 130$, the null hypothesis of no difference between R and P cannot be rejected. The p-

value is appropriately adjusted for the stopping boundary.[3] The plot should vary around a line from the origin with slope $\beta = -log(HR)$, with HR the hazard ratio for treatment effect (R:P).

Power at various true HRs is given in column 2 of the table below. In column 3 are given the numbers of events expected upon reaching termination. Also listed are the associated expected trial durations. (See the Technical Note in Section 3.5.) Computations of duration assume recruitment of 1440 patients in 33 months, a cumulative event rate in the placebo arm of 25% at 2 years, and losses to follow-up occurring at an annual rate of 5%.

2.5 Trial Power and Duration[@]

True HR	Power (in %)*	Events** at Termination	Estimated Trial Duration (months)				Upper Quartile[##]	Total Duration[##]
			Recruit	Add'l F-u	Extra[#]	Total		
1.00	2.5	191	27	0	1	28	250	35
0.90	17	248	33	2	1	36	320	42
0.85	35	270	33	4	1	38	346	45
0.80	58	273	33	5	1	39	350	46
0.75	**80**	**253**	**33**	**4**	**1**	**38**	**322**	**44**
0.70	93	271	33	1	1	35	271	40
0.65	99	177	39	0	1	30	220	36

[@] All durations are measured from July 1, 2012. Due to the delayed start-up, and according to the revised Accrual Milestones, we act as if recruitment is carried out evenly over the 33 month period July 2012 through March 2015, with 130 patients every 3 months.

* probability of a positive trial

** in the two arms together. The actual number is random and highly variable, depending on how the trial develops; what is tabled is the expected number to reach a stopping boundary (that is, averaged over many such trials under the same conditions). However, any particular trial can require as many as 130*4=520 events ($V = 130$), although it is highly unlikely – with probability < 0.001.

[#] The extra month is allowance for endpoint adjudication time. Again, total duration is the expected duration under the stated conditions, but actual duration may vary. `Duration' is measured from the time enrollment is underway at most centers until the trial ends.

[##] The last two columns give the upper quartile of the number of events at termination and the corresponding trial duration.

As seen in the table, under the stated assumptions, whatever the true hazard ratio, the trial is expected to require at most 38 months once the speed up of randomization begins, with the trial estimated to end in August 2015 leaving time to allow for closeout and analysis; 44 months duration would still allow closeout and analysis within a one-year no-cost extension. However, these computations are based on the very conservative assumption of a 25% 2-year event rate in the placebo arm whereas the revised eligibility will likely lead to a higher event rate since higher proportion of patients with prior VT/VF will be expected. Any increase in the event rate will speed up termination of the trial. Also, a smaller HR will shorten duration of the trial.

3 Analysis

3.1 Primary Analysis

At the end of the trial, a p-value for the primary hypothesis, an estimate of the true hazard ratio for treatment effect, and 95% confidence limits for the true hazard ratio will be determined, by methods adjusted to the sequential stopping rule.[3] The primary analysis will be a statistical test of treatment effect based on a Cox proportional-hazards regression analysis stratified by enrolling center – as used in the sequential monitoring – with six additional baseline covariates: ejection fraction, creatinine, age (all three numerical), ischemic status (binary), antiarrhythmic medication at enrollment (binary) and a 3-level variable identifying the following groups: - ICD and no history of VT/VF/cardiac arrest at enrollment; - CRT-D and no history of VT/VF/Cardiac Arrest at enrollment; - history of VT/VF/cardiac arrest, whether prior to implantation of device or afterwards. Device type is not expected to have an effect on risk in this third subgroup. This last subgroup does not distinguish between device types as they are not expected to have an effect on risk once VT/VF or a cardiac arrest has been experienced. These risk factors were chosen as being those found to be relevant in corresponding (unpublished) analyses of data from the MADIT-II and MADIT-CRT trials. Computations will be done by software developed at the University of Rochester, and will be confirmed by use of PEST software.[2] Some additional events will likely be reported after formal trial termination (events that occurred prior to termination), and these will be incorporated in the final adjusted p-value, hazard ratio and confidence limits computations.[4]

Multivariate imputation by chained equations[5,6] will be used to impute missing covariate data needed for the primary analysis. This will be used in the final analysis and in the monthly sequential monitoring. Predictive mean matching will be used to impute ejection fraction, creatinine, and age. A logistic regression model will be used to impute ischemic status and antiarrhythmic medication at enrollment. A multinomial logit model will be used to impute the 3-level variable incorporating type of device and history of VT/VF/cardiac arrest at enrollment. Each of these covariates will be imputed based on a model including the other five baseline covariates used in the primary analysis. Each imputed data set will be generated using 50 cycles (iterations), and this will be repeated 100 times to produce 100 multiply imputed data sets. However, when a stopping boundary is approached in the monthly monitoring, the number of multiply imputed data sets will be increased repeatedly to ensure that the Monte Carlo error is negligible, and hence to be confident whether a stopping boundary has been crossed. The same approach will be used in the final analysis.

The final primary analysis will be conducted after all events have been adjudicated. It is possible that neither the upper nor the lower stopping boundary will be reached before this final analysis. However, the path of the score-test statistic Z (when plotted against its variance V) will reach a terminal point at this final analysis, since no additional events or follow-up will occur. Essentially, the original triangular design would be truncated with a vertical boundary at the variance for the final analysis, say at $V = t^*$. If this truncation occurs, the chance of hitting the upper or lower boundary, and hence the significance level, will be reduced. To maintain a 5% significance level in the event of trial truncation, the rejection region will be expanded to include a small portion of the vertical truncation boundary. Specifically, the rejection region will include the upper boundary from $V = 0$ to $V = t^*$ and the vertical boundary from $Z = 14.8153 + 0.101103 \times t^*$ (upper boundary at $V = t^*$) down to $Z = c$, where c is chosen such that the chance of hitting this rejection region under the null hypothesis is 5%.[7]

3.2 Validation and Assumptions

Computations of p-values are broadly valid, but estimation of HRs presume a (near) constant hazard ratio. Hence, interpretation of the HR estimates in the primary analysis depends on validating the proportional hazards assumption. This will be done by computing HRs by both 3- and 6-month intervals, with tests for differences among the time-specific HRs.

3.3 Sensitivity Analysis

- Baseline covariate balance between arms: The primary analysis will be repeated including each (one at a time) baseline covariate in the regression model that is out of balance between arms.
- Evaluation of various combinations of the 3 components of the composite endpoint: A competing-risk analysis for separate components of the endpoint, determining a hazard ratio R:P for each, will be carried out,[8] in particular, for VT/VF requiring ATP, for ICD shock, for death, and for the first of ICD shock and death. The last of these provides an evaluation of the original composite endpoint without inclusion of the VT/VF requiring ATP component. Power for the others is not predicted, but power for the latter – namely ICD shock /death – is estimated to be 80% at a hazard ratio of 0.72.

3.4 Treatment Interactions

The primary analysis will be repeated (without adjustment for the stopping rule), adding each of a pre-specified list of covariates (if not already in the regression model), one at a time, and their interaction with treatment arm to the regression model, and tests for interaction carried out. This will identify, to the extent feasible, different treatment effects of Ranolazine across subgroups identified by the covariate.

- CRT versus ICD groups
- Primary versus secondary prevention groups
- Ischemic versus non-ischemic groups
- Females versus males
- Older versus younger
- BNP> versus <= median
- Diabetes mellitus (yes versus no)
- Antiarrhythmic medication at baseline
- Large centers versus small centers

3.5 Technical Note

The formula for the number N of events expected in a single arm of the trial, assuming recruitment of n patients to the arm in m months and then an additional f months of potential active follow-up, a monthly rate b of losses, and a monthly rate r of endpoint events, may be shown to be

$N = n*(r/s)*\{1 - [\exp(-s*f) - \exp(-s*d)]/(m*s)\}$ with $s = r + b$ and $d = m + f$.

(Derivation involves integration over exponentially distributed endpoints and losses and over uniformly distributed randomization.) Computing N using the rate r for the P arm, and again with r replaced by HR*r for the R arm, and adding, gives the number of events expected in a trial of duration d months. For the P arm, r = [-log(1-0.22)]/24 = 0.0103526; and b = [-log(1-0.05)]/12 = 0.004274, n = 720 and m = 27. Carrying out this computation for a list of f values, and various values of HR, provides corresponding

pairs of values for total duration *d* and the corresponding total number of events. This analysis ignores any potential effect of including six risk factors in the regression analysis, and is hence conservative.
PEST software (or other sequential trial software) can provide the number of events expected at termination of a sequential trial, in column 3 of the table above. Using the list of event-duration pairs, this leads to corresponding total duration (after adding in the additional month for adjudication time) for each row of the table. The number 720 of patients in each arm was found by trial and error as that value *n* needed to assure satisfactory total duration times. The last two columns of the table were similarly determined.

3.6 Secondary Analysis

The primary analyses with primary endpoints are planned with 80% power to detect specified differences between the ranolazine and placebo arms. Secondary aims will likewise require power of 80% (or more), where possible.

The first three secondary aims (see Section 1.2) are similar to the primary aim except with different composite endpoints. Hence, analysis of each will be similar to that for the primary aim, except that no adjustment for the sequential stopping rule will be feasible (as stopping is based on the primary composite endpoint).

1. Secondary aim #1, as mentioned above, power for ICD shock /death is estimated to be 80% at a hazard ratio of 0.72.
2. Secondary aim #2, in which hospitalizations for cardiac causes or death is the secondary endpoint, it is expected that at least 30% of patients in the placebo arm will reach this endpoint, leading to considerably more power than for the primary endpoint, namely 80% at a 25% reduction in the ongoing risk.
3. Second aim #3 addresses the effect of ranolazine on the composite endpoint consisting of CHF hospitalization or death. We assume that a 2-year probability of this endpoint in the placebo arm should reach at least 20%, resulting in power exceeding 80% for detecting a 30% reduction in the ongoing risk of this endpoint.
4. Secondary aim #4 is about repeated hospitalization for cardiac causes; for it, an Andersen-Gill regression analysis (comparable to Cox analysis for a first event), but with death as a competing risk,[8,9] will be carried out to assess any difference in ongoing risk of repeated therapy in the two arms of the trial. Power at comparable risk reductions should exceed that for secondary aim #2; this is confirmed by analysis of comparable subsets in the MADIT II data.
5. Secondary aim #5 is about repeated ICD therapies; for it, again (see aim #4 above) an Andersen-Gill regression will be carried out to assess any difference in ongoing risk of repeated therapy in the two arms of the trial. Power at comparable risk reductions should exceed that for the primary endpoint; this is supported by results from the azimilide SHIELD trial.
6. Secondary aim #6 is about inappropriate shocks. For first inappropriate shocks, a Cox regression analysis will be done while for repeated shock episodes, an Anderson-Gill analysis will be done; each of these will need to treat death as a competing risk.[8,9] The rate of first inappropriate shocks at 2 years is expected to be 16% (or more), allowing 80% power to detect a HR of 0.70. Power to detect similar effects for repeated shock episodes will be greater.
7. Secondary aim #7 is about quality of life (QoL – see Protocol Section D.8). For each patient having both baseline and 2-year QoL data, the change in QoL from baseline to 2 years will be determined and averaged over patients in each arm. Comparison of mean changes in the two arms will be evaluated by a t-test. We expect at least 500 patients in each arm to have the needed 2-year QoL

data (at least those recruited during the first 13 months). Earlier experience with the Kansas City Cardiomyopathy Questionnaire suggests a standard deviation for a single change in scores to be approximately 20. This should allow 80% power to detect a difference between arms in mean scores of 2.5 points and 90% power to detect difference of 3.0 points. Mean scores are expected to be in the neighborhood of 30 to 40. For the 6MWT, we will compare distance achieved by patients at baseline and at 2-year follow-ups, again limited to those with both 6MWTs, as measure of physical functioning. The primary analyses will be focused on the 2-year time-point and similarly to the quality of life analyses, we expect that at least 500 patients will have these tests performed. Based on prior studies, a standard deviation for a single change in the distance of the 6MWT is expected to be approximately 50 meters. This should allow 80% power to detect a difference between arms in mean distance of 6.5 meters and 90% power to detect a 7.5-meter difference in mean distance of 350 meters.

8. Secondary aim #8 is about safety of ranolazine therapy; each type of safety issue will be individually analyzed by use of ICD interrogation data, and summary statistics will be compiled.

4 References

1. Moss AJ, Zareba W, Hall WJ, Klein H, Wilber DJ, Cannom DS, Daubert JP, Higgins SL, Brown MW, Andrews ML; Multicenter Automatic Defibrillator Implantation Trial II Investigators. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. N Engl J Med. 2002;346:877-83. PMID: 11907286.
2. Brunier HC, Whitehead J. PEST 3.0 Operating Manual . The University of Reading, Reading, UK, 1993.
3. Whitehead J. The Design and Analysis of Sequential Clinical Trials. Rev'd 2nd ed., John Wiley, New York, 1997.
4. Hall WJ, Liu A. Sequential tests and estimators after overrunning based on maximum likelihood ordering. 2002; Biometrika 89:699-707.
5. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. 2011; Statistics in Medicine 30: 377-399.
6. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. 2011; Journal of Statistical Software 45(3): 1-67.
7. Hall WJ. The distribution of Brownian motion on linear stopping boundaries. 1997; Sequential Analysis 16:345-352. Addendum in 1998; Sequential Analysis 17:123-124.
8. Lawless JF. Statistical Models and Methods for Lifetime Data. 2nd ed., John Wiley, New York, 2003.
9. Cook RJ, Lawless JF. The Statistical Analysis of Recurrent Events. Springer, New York, 2007.