

Endovascular Therapy Following Imaging
Evaluation for Ischemic Stroke 3

Statistical Analysis Plan

NCT02586415

September 8, 2017

Gregory W. Albers, MD, Principal Investigator
Stanford University
Stanford, California 94305

Memorandum

Date: July 19, 2017

To: The DEFUSE 3 SAP

From: [REDACTED]

RE: Interim Analysis Plan, as requested by the DEFUSE 3 DSMB

An early interim analysis was requested by NINDS, in response to external information generated from the DAWN trial, in order to make a determination as to whether DEFUSE 3 should continue in the original population or a subpopulation, or be terminated early.

The details of this proposed analysis were put forth in Version 2.0 of the DEFUSE 3 SAP. On June 23, via email communication from [REDACTED], a revision to that analysis plan was requested. The details of the plan, as requested by the DSMB, are outlined below; we have noted details which deviate from the plan outlined in the SAP. This information is documented in a memo to the SAP, rather than a revision to the SAP, in order to keep the study team blinded.

For this early interim analysis, the overall null hypothesis for the primary endpoint will be tested using the Wilcoxon rank sum statistic and asymptotic pvalue, generated via the PROC NPAR1WAY procedure in SAS. The 90-day mRS for subjects who have not yet reached the end of the study protocol will be imputed with their respective 30-day mRS. Subjects with no 30-day or 90-day mRS scores will be excluded. This is as specified in Version 2.0 of the DEFUSE 3 SAP. The SAP stipulates a one-sided level of significance of 0.023 for this analysis; however, the DSMB will meet to decide on decision rules prior to reviewing the results, and their decision rules are presumed to override the level of significance stated in the SAP.

The DSMB requested that the above analysis be repeated in the subgroup of DEFUSE 3 subjects who are considered DAWN-eligible. The DSMB will decide whether to review the subgroup analysis at the time of the meeting.

Per DSMB request, the power of these analyses at the available sample size will be calculated, as will the nominal threshold for declaring efficacy or harm if this was a typical unplanned interim analysis.

The DSMB requested an additional evaluation of harm associated with the endovascular therapy. This analysis plan is based on telephone consultation with [REDACTED] (on July 7) and email communication from [REDACTED] (on July 13). Briefly, the probability of a 'harmful' event under each treatment group ($\pi_{\text{endovascular}}$ and $\pi_{\text{medical management}}$) will be considered as random variables with prior probability distribution Beta (1,1), representing a uniform distribution. The posterior distributions are also represented by Beta, where the parameters are defined according to the number of subjects (and the number of harmful events) in each treatment arm. The distribution of the difference between $\pi_{\text{endovascular}}$ and $\pi_{\text{medical management}}$ will then be estimated via simulation. This process will be repeated for two definitions of harmful event: mRS 4-6 and mRS 3-6. The analysis will be conducted in the full cohort as well as in the DAWN-eligible subgroup.

The DSMB further requested that the absolute treatment effect, according to dichotomy defined by mRS 0-2 vs 3-6, and its corresponding 95% confidence intervals, also be presented. Again, the intervals will be constructed for the full cohort as well as the DAWN-eligible subgroup.

The semi-annual report for the full DEFUSE 3 cohort was emailed to Peter Gilbert on July 19. The semi-annual report based on the DAWN-eligible subgroup, as well as the requested analyses (in the overall cohort and in the DAWN-eligible subgroup) , will be brought to the meeting by the unblinded study statisticians. The individual analysis documents will be held in separate, signed and sealed envelopes; after agreeing on the decision rules for stopping the study, the unblinded statisticians will provide the DSMB with the relevant results for the full cohort. The DSMB will determine which of the ancillary analyses it wishes to see at that time.

Memorandum

Date: July 24, 2017
To: The DEFUSE 3 SAP
From: [REDACTED]
RE: the DSMB review of interim analysis results

An early interim analysis was requested by NINDS, in response to external information generated from the DAWN trial, in order to make a determination as to whether DEFUSE 3 should continue in the original population or a subpopulation, or be terminated early.

The details of this proposed analysis were put forth in Version 2.0 of the DEFUSE 3 SAP. On June 23, via email communication from [REDACTED], a revision to that analysis plan was requested. The details of the plan, as requested by the DSMB, are outlined below; we have noted details which deviate from the plan outlined in the SAP. This information is documented in a memo to the SAP, rather than a revision to the SAP, in order to keep the study team blinded.

For this early interim analysis, the overall null hypothesis for the primary endpoint was tested using the Wilcoxon rank sum statistic and asymptotic pvalue, generated via the PROC NPAR1WAY procedure in SAS. The 90-day mRS for subjects who have not yet reached the end of the study protocol was imputed with their respective 30-day mRS. Subjects with no 30-day or 90-day mRS scores were excluded. This is as specified in Version 2.0 of the DEFUSE 3 SAP. The SAP stipulated a one-sided level of significance of 0.023 for this analysis; however, the DSMB met to decide on decision rules prior to reviewing the results, and their decision rules override the level of significance stated in the SAP.

Per DSMB request, the nominal threshold for declaring efficacy or harm if this was a typical unplanned interim analysis was provided. Two boundaries were calculated under a general design, according to a maximum sample size of 476 subjects (as the original maximum for the adaptive design) and according to a maximum sample size of 376 subjects (the sample size required for the non-adaptive design). The DSMB decided to proceed under the latter (boundary provided in the table below), and that the analysis should consider a two-sided alternative. Dr. [REDACTED] then emailed the board members the password for the corresponding document.

Look	Information Fraction	2-sided Efficacy Boundary
		Z-scale
Interim	0.48	± 3.018
Final	1.0	± 1.967

The DSMB then requested to see the two-sided ordinal analysis results for the DAWN-eligible subgroup, and Dr. [REDACTED] emailed the board members the password for the corresponding document.

The DSMB then reviewed the closed semi-annual report for the full cohort.

DEFUSE 3

STATISTICAL ANALYSIS PLAN

Prepared by

Greg Albers, MD

[REDACTED], PhD

Maarten Lansberg, MD

[REDACTED], PhD

[REDACTED], Ph.D.

[REDACTED], Ph.D.

Stanford University

and

DEFUSE 3 National Data Management Center

Data Coordination Unit

Department of Public Health Sciences

Medical University of South Carolina

Charleston, SC 29425

Version 2.0, June 3, 2017 (Added changes to address early analysis as described in Section 2)

Version 3.0, August 24, 2017 (Added clarifications for programming purposes prior to unblinding of Stanford Study Team)

Version 3.1, September 8, 2017 (Added clarifications for programming purposes prior to unblinding of Stanford Study Team)

Table of Contents

1. LIST OF ABBREVIATIONS	3
2. STATISTICAL ANALYSIS PLAN, STATISTICAL REPORTS AND EARLY INTERIM ANALYSIS	3
3. STUDY OBJECTIVES	5
3.1. EFFICACY	5
3.2. SAFETY	5
4. STUDY DESIGN	5
5. DEFINITION OF ELIGIBLE POPULATION AND CHOICE OF ANALYSIS SET	5
5.1. ELIGIBLE POPULATION	5
5.2. ADAPTIVE DESIGN SAMPLE	6
5.3. SAFETY SAMPLE	6
6. RANDOMIZATION	6
7. BLINDING	7
8. MISSING DATA	7
9. PRIMARY EFFICACY ANALYSIS	8
9.1. PRIMARY OUTCOME	8
9.2. IMPACT OF ADAPTIVE DESIGN ON THE SAMPLE FOR THE PRIMARY ANALYSIS	8
9.3. INTERIM AND FINAL STATISTICAL ANALYSES	9
9.4. REPORTING OF PRIMARY RESULTS	11
9.5. ESTIMATION OF P-VALUES, EFFECT SIZE ESTIMATES, AND CIs	12
10. SAMPLE SIZE DETERMINATION FOR PRIMARY EFFICACY ANALYSIS	12
11. EXPLORATORY ANALYSES OF THE PRIMARY OUTCOME AFTER TRIAL COMPLETION	13
11.1. ANALYSIS ADJUSTING FOR COVARIATES	13
11.2. ANALYSIS UNDER THE AS-TREATED PRINCIPLE	15
11.2.1. GROUP ASSIGNMENT UNDER THE “AS-TREATED” PRINCIPLE	15
12. ANALYSES OF SECONDARY EFFICACY OUTCOMES	15
12.1. ANALYSES OF SECONDARY CLINICAL EFFICACY OUTCOMES	15
12.2. ANALYSES OF FUNCTIONAL INDEPENDENCE IN SUBGROUPS	15
12.3. ANALYSES OF IMAGING EFFICACY OUTCOMES	16
13. SAFETY ANALYSES	17
13.1. MONITORING OF DEATHS AND SICH	17
13.2. SUMMARY OF ADVERSE EVENTS (AEs) AND SERIOUS ADVERSE EVENTS (SAEs)	17
14. COORDINATION BETWEEN STANFORD AND MUSC STATISTICAL TEAMS	18
15. REFERENCES	18

1. List of abbreviations

AE	adverse event
ASPECTS	Alberta Stroke Program Early CT Score
CRF	case report form
CT	computer tomography
CTA	computer tomography angiography
DCU	Data Coordination Unit at the Medical University of South Carolina
DCR	Data Clarification Request
DSMB	Data and Safety Monitoring Board
EC	Executive Committee
ICA	internal carotid artery
ICH	intracranial hemorrhage
IMM	independent medical monitor
ITT	intent-to-treat
IV	intravenous
LTFU	lost to follow up
MCA	middle cerebral artery
MRI	magnetic resonance imaging
MRA	magnetic resonance angiography
mRS	modified Rankin Scale
NDMC	National Data Management Center
NIHSS	National Institutes of Health Stroke Scale score
NINDS	National Institute of Neurological Disorders and Stroke
OR	odds ratio
RR	relative risk
rt-PA	recombinant tissue plasminogen activator
SAE	serious adverse event
SAP	statistical analysis plan
sICH	symptomatic intracranial hemorrhage
TICI	thrombolysis in cerebral infarction

2. Statistical analysis plan, statistical reports and early interim analysis

This statistical analysis plan (SAP) was modified on June 3, 2017 to accommodate significant external events that occurred in May 2017. On May 16, 2017, the results of DAWN, a clinical trial that enrolled similar patients and studied a similar intervention as DEFUSE 3, were presented at an international meeting. The results of this study demonstrated a substantial clinical benefit of endovascular therapy over medical therapy. Based on the results of the DAWN study, on May 24, 2017, the DEFUSE 3 Central IRB requested that enrollment in DEFUSE 3 be halted; at that time, 182 patients had been randomized. On May 26, 2017, the DSMB and the DEFUSE 3 Executive Committee both recommended that the trial be halted. The NINDS instead requested an early data analysis and revision of the DEFUSE 3 Statistical Analysis Plan (SAP) to provide for the option of continuing the study if the early interim analysis did not cross the stopping boundary for efficacy. Based on this request, the SAP has been modified to provide for this unplanned early analysis. This analysis will be conducted as follows:

Upon database freeze, the DEFUSE 3 unblinded Statistician (Dr. [REDACTED]) will perform a test of the overall null hypothesis for the primary endpoint (mRS shift analysis) in the full DEFUSE 3 sample at the one-sided Type 1 error probability of 0.023 (additional details of how this analysis will be performed are presented in section 9.4). For the early interim analysis, we will impute the

90-day mRS for subjects who have not yet reached the end of the study protocol with their respective 30-day mRS. Subjects with no 30-day or 90-day mRS scores will be excluded. The DSMB will be provided with the results of this analysis and would likely recommend stopping the study if the result is significant at the one-sided 0.023 level. Subsequently, upon submission of the last-randomized subject's 90-day outcome data into the WebDCU™ (anticipated in mid-August 2017) and the database lock, the analysis will be repeated at the one-sided alpha level of 0.023. The manuscript for the DEFUSE 3 Trial results will reflect this second analysis.

If the initial analysis result (i.e., based on the frozen, not locked, database) is not significant at the one-sided alpha level of 0.023, the DSMB will review the result and may request any additional analyses in any subgroups that they desire in order to make recommendations to NINDS. Following the DSMB review and recommendation, the NINDS may make the decision to continue the study with recruitment of the DEFUSE 3 eligible patients per the current eligibility criteria, and the study may continue per the current protocol. In this case, the final primary outcome analysis will be tested at the one-sided alpha=0.001 (which will be referred to as "alpha2"). If the decision is made to terminate the study, upon submission of the last-randomized subject's 90-day outcome data into the WebDCU™ (anticipated in mid-August 2017) and the database lock, the analysis will be repeated at the one-sided alpha level of 0.023.

The NINDS may also direct the study team to continue the study in a subset of the current DEFUSE 3 eligible patient population (such as DAWN-ineligible patients). If this is deemed feasible, then the final analysis will include all patients enrolled (both before and after the subset selection was imposed) and will be evaluated with a one-sided alpha of 0.001 (which will be referred to as alpha3).

The rationale for recommending the extreme split of the alpha is that:

- (a) given the conditions currently imposed by the CIRB (i.e., recruit only DAWN-ineligible patients), continuing the DEFUSE 3 study with the current eligibility criteria is unlikely to be feasible; and
- (b) the sites have reported significant concerns regarding recruiting the DAWN-ineligible patients due to its complex definition.

To date, the DEFUSE 3 Executive Committee has proposed and the DSMB had effectively recommended $\alpha_2 + \alpha_3 = 0$, since they do not envision any circumstance where continuation of DEFUSE 3 in its current form is feasible. If a continuation is requested by the NINDS, the sample size will need to be reassessed based on the desired effect in the modified study population and the one-sided α_2/α_3 required for efficacy. This evaluation will likely result in increased total sample size for the study.

This proposed plan supersedes the prior SAP. No further interim analyses for adaptation will be conducted, and Sections 9.2-9.4 of the prior SAP are now deleted. Other changes include: Subgroup analysis related to time from symptom onset to randomization and baseline core volume (Section 12.2 of the prior SAP) define subgroups based on medians, rather than "cutpoint from adaptive design". An additional pre-specified exploratory analysis has been included that compares the primary and secondary efficacy and safety endpoints in DAWN-eligible vs DAWN-ineligible patients (see section 11.1). The remainder of the SAP will be followed as pre-specified prior to this revision.

This document provides the details of the statistical analyses planned for the DEFUSE 3 Trial, including the original interim analyses for efficacy, futility, and subgroup selection, and the revision to reflect the early interim analysis for overwhelming efficacy. In addition, it discusses

the statistical issues relevant to these analyses (e.g., sample data to be used, imputation of missing data, adjustments for multiplicity, etc.).

The NDMC generates DSMB Reports semiannually. Each semiannual report provides cumulative summary statistics on enrollment; subject status in the study (e.g., number completed 30 and 90 day assessments); baseline characteristics; protocol violations; safety data, including AEs and SAEs by AE code and relatedness to the study intervention; and data management/quality information (e.g., timeliness and completeness of data entry by the clinical centers via the StrokeNet WebDCU™ Website; number of DCRs generated and resolved). These statistics are reported by treatment group.

3. Study Objectives

3.1. Efficacy

The primary objective of the DEFUSE 3 Trial is to determine if ischemic stroke subjects treated in the 6-16 hour time-window with endovascular therapy plus medical management have more favorable functional outcomes at 90 days, defined by mRS score, as compared to subjects treated with medical management alone.

For supportive evidence, the trial plans to evaluate the effectiveness of endovascular therapy plus medical management as compared to medical management alone by other clinical measures (e.g., mRS 0-2 outcomes at 90 days) and imaging data (e.g., proportion with reperfusion and infarct growth at 24 hours).

3.2. Safety

The safety of endovascular therapy plus medical management as compared to medical management alone is monitored and evaluated by deaths and incidence of sICH, and other SAEs.

4. Study Design

The study has a two-arm parallel design. Eligible subjects are randomized in a 1:1 ratio to endovascular therapy plus medical management or medical management alone. Each subject is followed for 3 months from randomization.

5. Definition of eligible population and choice of analysis set

5.1. Eligible Population

At the outset of the DEFUSE 3 trial, an eligible patient has an acute ischemic stroke, is 18-90 years of age, has an NIHSS of at least 6 and no more than 24, has no significant pre-stroke disability (pre-baseline mRS of 0-2), can undergo endovascular therapy between 6 and 16 hours of stroke onset, and has evidence of a large vessel occlusion and a large penumbra by neuro-imaging. ~~At one of the two interim analyses, the study inclusion criteria may be altered by the adaptive design (see Section 9).~~ The specific neuro-imaging criteria to qualify for randomization at the onset of the study are:

- ICA or MCA-M1 occlusion (carotid occlusions can be cervical or intracranial; with or without tandem MCA lesions) by MRA or CTA; AND
- Target Mismatch Profile on MRI or CT perfusion (ischemic core volume is <70 ml, mismatch ratio is ≥ 1.8 and mismatch volume is ≥ 15 ml as determined by RAPID software).

If perfusion imaging, the MRA, or the CTA is technically inadequate, alternative neuro-imaging inclusion criteria to qualify for randomization are:

A) If CTA (or MRA) is technically inadequate

- Tmax>6s perfusion deficit consistent with an ICA or MCA-M1 occlusion; AND
- Target Mismatch Profile (ischemic core volume is <70 ml, mismatch ratio is ≥ 1.8 and mismatch volume is ≥ 15 ml as determined by RAPID software).

B) If MRP is technically inadequate

- ICA or MCA-M1 occlusion (carotid occlusions can be cervical or intracranial; with or without tandem MCA lesions) by MRA (or CTA, if MRA is technically inadequate and a CTA was performed within 60 minutes prior to the MRI); AND
- DWI lesion volume <25 ml

C) If CTP is technically inadequate

- Patient can be screened with MRI and enrolled if neuroimaging criteria are met.

5.2. Adaptive design sample Analysis Sample

~~The adaptive design sample is the group selected by the adaptive design. It includes all randomized subjects (all six cells; Figure 1) if no subgroup is selected; it includes a subset of all randomized subjects if the adaptive design results in the selection of one of five possible subgroups (see Figure 1 and Section 9). The primary efficacy analysis will be conducted in the adaptive design sample (see Section 9). The analysis sample consists of all randomized patients and the primary analysis will be conducted per the ITT principle.~~

5.3. Safety sample

The safety sample includes all randomized subjects. Thus, the safety sample is the same, regardless of whether a subgroup is selected to continue enrollment ~~by the adaptive design.~~ See Section 13 for details of the safety analyses.

6. Randomization

Randomization takes place centrally via the DEFUSE 3 Trial WebDCU™ website. The randomization scheme is the combination of minimization and the biased coin method and is never deterministic. A dynamic stratification system will ensure well-balanced subgroups. The randomization algorithm will employ biased-coin minimization and the variance method with stratification weights.¹ The strategy is to balance treatment assignment along the marginal distribution of each stratification factor. The stratification factors used and their hierarchy will be: 1) ischemic core volume, 2) age, 3) time from symptom onset to enrollment, 4) NIHSS score and 5) study site. When a new patient is enrolled, the site will enter the stratification factor values into the eCRF (electronic case report form) on WebDCU™. The dynamic randomization

algorithm will determine an imbalance measure for each treatment group. The treatment group associated with the smallest imbalance measure will receive the largest probability of assignment in the biased-coin randomization. The randomization algorithm will be programmed into the WebDCU™ and validated using test samples by the NDMC. The detailed randomization scheme, including biased coin acceptance region and stratification weights, and source codes are provided in the Randomization Plan document. (Appendix 1)

7. Blinding

The acute treatment phase of the study is conducted in an open-label manner. However, study investigators who are not directly involved with acute treatment of the subject and who are blinded to treatment assignment will conduct all 30 and 90-day outcome assessments. To maintain blinding of the assessor, subjects are instructed not to discuss their initial hospitalization and treatment with the assessor.

In cases where an unblinded assessor performed the 30- or 90-day assessment (the CRF will capture if the assessor was blinded; this variable will be self-reported by the assessor), such occurrence will be marked as a protocol violation and presented in the final study report; nevertheless, the submitted data are used in the analysis.

8. Handling of Missing Data at the Final Analysis

Based on previous experiences with acute stroke trials, it is anticipated that there will be minimal loss to follow up for the 90-day assessment of the primary outcome. In the IMS I Trial, only 1 of 80 (1.25%) subjects were LTFU. In the IMS II Pilot Study, 2 of 73 (2.7%) were LTFU. In the IMS III Study, 27 of 656 (4.1%) subjects were LTFU. In SWIFT-PRIME, 4 of 191 (2.1%) were LTFU. In Fast-Mag, out of 1700 patients, 4 patients (0.2%) did not have at least day 30 follow-up.

~~In DEFUSE 3, at each analysis stage, the definitive sample for the boundary crossing analysis will consist of the first N consecutively recruited subjects, where N is the design specified sample size at each interim stage or the total sample at the final stage, who are in the selected subgroup if one has been chosen. All effort is put forth to ensure near complete follow-up, in particular with the assessment of the primary outcome (mRS at 90 days), death (mRS=6), and stroke recurrence. If the primary outcome (mRS at 90 days) cannot be assessed in the clinic, it will instead be obtained by phone using a structured interview. The final analysis will be conducted when all subjects in the analysis sample have reached the upper limit of the window for the 90-day outcome (120 days after randomization), and when the coordinating center believes it has exhausted all reasonable efforts to obtain outcome data collected within the window but delayed in data entry.~~

Missing 90-day mRS data (no mRS available within a 60-120 day window) will be handled by a hybrid approach: if the 30-day mRS is available, it will be “carried forward”, if not, the 90-day mRS will be “multiply imputed” (multiple imputation) using 'age' and 'NIHSS score at hospital discharge' as predictor variables.

All final analyses described in the SAP will use the multiple imputation data if applicable. ~~Specifically, the standardized, multiple imputation adjusted Wilcoxon rank sum statistic will be calculated. The Kullback-Leibler score, used in the~~

~~adaptive design for subgroup selection, will also be estimated using multiple imputation data. (see section 9.2)~~

~~The multiple imputation model will be based on all data accumulated in the study, and all missing data will be imputed based on the final model. If, at any stage, the adaptive design specifies testing of the null in a subgroup, the standardized, multiple imputation adjusted Wilcoxon-Mann-Whitney test statistic will be calculated in that subgroup without refitting the multiple imputation model or changing the imputations.~~

9. Primary efficacy analysis

~~Regardless of whether or not a subgroup is selected. All efficacy outcome measures are analyzed under the ITT principle. Under this principle, each subject is analyzed according to the treatment group to which they were randomly assigned. Definition of the sample included in the primary efficacy analysis (the adaptive design sample) is listed below in Section 9.3.~~

9.1. Primary outcome

The primary efficacy outcome measure is the mRS score at 90 days from randomization. Missing outcome is imputed according to Section 8.

9.2. Impact of adaptive design on the sample for the primary analysis

~~An adaptive trial design, developed for DEFUSE 3, will allow the study to test the primary efficacy hypothesis in a subpopulation (the adaptive design sample) if an interim or final analysis indicates futility in the overall population.² The design is based on closed testing theory and the group sequential methods for the Generalized Likelihood Ratio (GLR) statistic developed by Lai and Shih.³ The adaptive design was chosen because there is strong preliminary data suggesting that the effect of endovascular treatment is modified by two baseline variables: ischemic core lesion size and time to treatment. The way the adaptive design takes advantage of these biological assumptions (when they are true) is by reallocation of future accrual to the subgroup with the best prospects for showing efficacy. Specifically, if a subgroup is chosen at an interim analysis, subsequent enrollment is limited to patients in that subgroup. As a result, this subgroup will become larger than it would have been in the absence of the adaptive design.~~

The criterion for deciding which subgroup has the best chance of showing a benefit from endovascular therapy combines both the estimated size of the effect in the subgroup and the sample size of the subgroup. The GLR statistic (Kullback-Leibler criterion) is used to identify this subgroup because it optimally balances those two criteria. It selects the subgroup that has the best chance of showing an effect because it has an apparently large effect and is also of substantial size (note there are 5 subgroups of increasingly larger size, **Figure 1**). The adaptive design employs two biologically based assumptions to limit the inflation of sample size: a monotonicity / contiguity assumption and an assumption that the effect is largest in the patients with the smallest volumes and the shortest time to randomization (cell C_{11} in **Figure 1**). The cut-points of the categories (cells) will be determined just prior to the first interim analysis ($n=200$), blinded to subjects' treatment allocations and outcomes, based on the distribution of subjects across the two dimensions (lesion volume and time to treatment) to yield six categories (cells) of approximately equal number of subjects.

		Time (hrs)	
		<10	10-16
Core lesion volume (ml)	≤ 20	C_{11}	C_{21}
	21-50	C_{12}	C_{22}
	51-70	C_{13}	C_{23}

Figure 1. The cohort is stratified according to core lesion volume and time to randomization. Cut points in the figure serve as examples. Exact cut points of the stratification will be determined, blinded to treatment allocation and outcome, based on the distribution of subjects across the core and time variables, just prior to the first interim analysis. Based on the results of the 1st interim analysis, enrollment will continue in all 6 cells or the study entry criteria will adapt and enrollment will be limited to one of 5 sub-groups (C_{11} , C_{11+21} , $C_{11+21+12}$, $C_{11+21+12+22}$, or $C_{11+21+12+22+13}$).

9.3. Interim and final statistical analyses

The primary endpoint is the distribution of scores on the modified Rankin Scale (mRS) at day 90. We will test the primary efficacy and futility hypotheses at the interim and final analysis using the generalized likelihood ratio (GLR) test, based on whether the usual normal approximation to the Wilcoxon-Mann-Whitney test statistic crosses a futility or efficacy boundary at interim or final analysis. The primary analysis will be conducted in the adaptive design sample (see section 5.2 for definition), according to the intention to treat principle, adjusted for the adaptive design, and unadjusted for covariates. See Appendix 2 for detailed specification of the interim and final calculations.

The efficacy bounds at interim and final analysis are set to control the overall (one-sided) Type I error rate at 2.5%. There are three group sequential boundaries: an interim futility boundary b_f , an interim efficacy boundary b_e , and a final efficacy bound c , which are fixed before the first interim analysis. At each of the two interim analyses, the futility bound b_f is used to decide if the study should continue recruitment in the overall group, shift accrual and testing to a subgroup, or stop in its entirety. The futility boundary takes into account the fact that the maximum analyzed sample size is a random variable that is no larger than the fixed maximum number of subjects randomized ($n=476$). Because subgroup selection reduces the maximum number of subjects available for analysis at completion of the study, this method effectively allows an easier futility stop after subgroup selection. This setup replaces conditional power analyses with an automatic and more efficient adjustment of boundaries.

If the patients are equally distributed across the six cells, the efficacy boundary at the interim analyses will be 2.62, the efficacy boundary at the final analysis will be 2.61, and the futility boundary at the interim analyses will be 1.88. If the patients are not distributed equally, these boundaries will be adapted slightly.²

First interim analysis (based on primary outcome data obtained from the first 200 consecutively randomized subjects)

The null hypothesis is tested in the entire subject population, and, depending on the results:

1. If neither efficacy nor futility bound is crossed, the trial continues with enrollment in the overall population to the 2nd-interim analysis.
2. If the efficacy bound is crossed, the trial stops and efficacy is declared in the overall population.
3. If the futility bound is crossed, the optimal subgroup is selected based on the Kullback-Leibler criterion and the null is tested in that subgroup.
 - 3.1. If neither bound is crossed, the trial will continue with enrollment limited to the selected subgroup.
 - 3.2. If the efficacy bound is crossed, the trial stops and efficacy is declared in the selected subgroup
 - 3.3. If the futility bound is crossed, the trial stops for futility.

Second interim analysis (after primary outcome data obtained from an additional 140 consecutively randomized subjects within the target population defined by the first interim analysis)

If, after the first interim analysis, the study proceeds with enrollment in the overall population (option 1, under first interim analysis), the testing at the 2nd-interim analysis is identical to the first interim and the decisions to stop or proceed with enrollment are identical to those outlined above under the first interim analysis.

If, after the first interim analysis, enrollment is limited to a selected subgroup (option 3, under first interim analysis), the second interim analysis is based on a test of the null hypothesis in the selected subgroup only and, depending on the results:

1. If neither bound is crossed, the trial continues to the final analysis with enrollment of additional subjects limited to the selected subgroup
2. If the efficacy bound is crossed, the trial stops and efficacy is declared in the selected subgroup
3. If the futility bound is crossed, the trial stops for futility. Note that there is no option for "second subgroup selection".

Final analysis (after primary outcome data obtained from an additional 136 consecutively randomized subjects within the target population defined by the second interim analysis)

If, after the second interim analysis, the study proceeds with enrollment in the overall population the null is tested in the overall population, and, depending on the results:

1. If the efficacy bound is crossed, endovascular therapy is declared efficacious in the overall population.
2. If the efficacy bound is not crossed, the optimal subgroup is selected and the null is tested in that group:
 - 2.1. If the efficacy bound is crossed, endovascular therapy is declared efficacious in that subgroup
 - 2.2. If the efficacy bound is not crossed, endovascular therapy will be declared of no benefit.

~~If enrollment after one of the interim analyses is limited to a selected subgroup, then at the final analysis the null will be tested in that subgroup only and efficacy or lack thereof will be declared as per options 2.1 and 2.2 above.~~

~~The computation of the test statistics and the Kullback-Leibler selection criterion are specified in Appendix 2. Calculation of the test statistics will be carried out in SAS by the NDMC. A more detailed description of the adaptive design is provided in Appendix 3.~~

9.4. Interim and Final statistical analysis

The overall null hypothesis for the primary endpoint will be tested using the Wilcoxon rank sum statistic and asymptotic pvalue generated via the NPAR1WAY procedure in SAS. Tied values will be assigned the average rank for the corresponding value, as is the default. The continuity correction will not be applied.

Based on external information generated from the DAWN trial, an early interim analysis was requested by NINDS in order to make a determination as to whether DEFUSE should continue in the original population or a subpopulation or be terminated early. For the early interim analysis, the efficacy hypothesis will be tested at the one-sided Type 1 error probability of 0.023. The 90-day mRS for subjects who have not yet reached the end of the study protocol will be imputed with their respective 30-day mRS. Subjects with no 30-day or 90-day mRS scores will be excluded. The DSMB will be provided with the results of this analysis and would likely recommend stopping the study if the result is significant at the one-sided 0.023 level. Subsequently, upon submission of the last-randomized subject's 90-day outcome data into the WebDCU™ (anticipated in mid-August 2017) and the database lock, the analysis will be repeated at the one-sided alpha level of 0.023. The manuscript for the DEFUSE 3 Trial results will ~~reflect~~ contain results from this second analysis executed on the final locked database.

If the interim analysis is not significant at the one-sided alpha level of 0.023, it is anticipated that the DSMB will review the results and may request any additional subgroup analyses in order to make recommendations to NINDS. Such requests will be documented in an addendum to the SAP. Following the DSMB review and recommendation, the NINDS may make the decision to continue the study with recruitment of the DEFUSE 3 eligible patients per the current eligibility criteria, and the study may continue per the current protocol. In this case, the final primary outcome analysis will be tested at the one-sided alpha=0.001 (which will be referred to as "alpha2"). If the decision is made to terminate the study, upon submission of the last-randomized subject's 90-day outcome data into the WebDCU™ (anticipated in mid-August 2017) and the database locked, the analysis will be repeated at the one-sided alpha level of 0.023.

The NINDS may also direct the study team to continue the study in a subset of the current DEFUSE 3 eligible patient population (such as DAWN-ineligible patients). If this is deemed feasible, then the final analysis will include all patients enrolled (both before and after the subset selection was imposed) and will be evaluated with a one-sided alpha of 0.001 (which will be referred to as alpha3).

The rationale for recommending the extreme split of the alpha is that:

(a) given the conditions currently imposed by the CIRB (i.e., recruit only DAWN-ineligible patients), continuing the DEFUSE 3 study with the current eligibility criteria is unlikely to be

feasible; and (b) the sites have reported significant concerns regarding recruiting the DAWN-ineligible patients due to its complex definition.

To date, the DEFUSE 3 Executive Committee has proposed and DSMB had effectively recommended $\alpha_2 + \alpha_3 = 0$, since they do not envision any circumstance where continuation of DEFUSE 3 in its current form is feasible. If a continuation is requested by the NINDS, a sample size will need to be reassessed based on the desired effect in the modified study population and the one-sided α_2/α_3 required for efficacy. This evaluation will likely result in increased total sample size for the study.

9.5. Reporting of primary results

The results of the study will be primarily expressed as whether or not an efficacy boundary was crossed at ~~either one of the two interim analyses or at the final analysis. The efficacy boundary is set, a priori, to guarantee less than a 2.5% one-sided error rate.~~ Crossing of the efficacy boundary will be considered evidence that endovascular therapy is beneficial, based on lower day-90 mRS scores in the endovascular group compared to controls.

9.6. Estimation of p-values, effect size estimates, and CIs after trial completion

The treatment effect will be ~~adjusted for study design and~~ expressed as

- The common odds ratio with its 95% confidence interval and p-value, calculated using a proportional odds model.
- The average number needed to treat for benefit (NNT), with its 95% confidence interval, where $NNT = 1 / (P_EndovascularSuperior - P_MedicalSuperior)$.⁴

10. Sample size determination for primary efficacy analysis

The sample size determination begins with a preliminary estimate of the effect size that is plausible to expect and which is also clinically meaningful. Fixing standard operating characteristics at 5% two-sided Type 1 error and 10% Type 2 error (90% power) leads to a sample size lower bound for a hypothetical fixed-sample trial with no adaptation and no interim analysis. We then adjust the sample size for the group-sequential modification and the subgroup adaptive design.

The projected overall effect of endovascular therapy is based on 1) the observed 90-day modified Rankin Scale outcomes in DEFUSE 2 of target mismatch patients treated >6hrs after symptom onset and 2) the assumption that early reperfusion will be achieved in 75% of the endovascular arm vs. 20% of the medical therapy arm.⁵⁻⁷ Using these data, we have projected the distributions on the mRS at 90 days for subjects in the endovascular and control arms of DEFUSE 3. (Table 1)

	mRS at day 90							Total
	0	1	2	3	4	5	6	
Endovascular group	18.0%	11.5%	19.6%	11.5%	16.4%	11.5%	11.5%	100%
Medical group	9.7%	7.9%	15.0%	17.7%	14.4%	17.7%	17.7%	100%

Table 1. Projected 90-day mRS distributions in DEFUSE 3

These distributions correspond to a standardized effect of 0.36 for the primary analysis. Based on these data, the fixed sample size for a non-adaptive design requires 376 patients (188 per arm) to have 90% power at an alpha of 5% (Wilcoxon-Mann-Whitney test); 100 patients were

added for the adaptive design to reach a maximum sample size of 476 for DEFUSE 3. The size of this increase is based on simulations and is selected to preserve the desired operating characteristics, while allowing shrinkage in effect size to 0.30, since the above estimate of 0.36 may be optimistic. The sample size of 476 is also the largest sample that can be accrued within budget and time limitations.

Simulations (n=5000) are used to compare the performance of a traditional fixed sample size design (fixed n=476) to the adaptive design (max randomized n=476) under the null and various alternative scenarios (**Table 2**). For the simulations the effect size is expressed as a standardized effect *in the disjoint cells*, which are cumulated to form the subgroups as described above, where a standardized effect of 0.3 corresponds to a conservative projected effect of endovascular therapy (anticipated effect 0.36; see above).

Sim.	Standardized effect in cells*						Average standard effect	Adaptive Design		Fixed-Design	
	C ₁₁	C ₂₁	C ₁₂	C ₂₂	C ₁₃	C ₂₃		Average No. randomized	Power	Number randomized	Power
#0	0	0	0	0	0	0	0	361	2.2%	476	2.5%
#1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	354	80%	476	89%
#2	0.5	0.4	0.3	0	0	0	0.2	400	86%	476	55%
#3	0.5	0.5	0	0	0	0	0.17	403	87%	476	41%

Table 2. Simulations comparing the adaptive and the fixed trial designs. *Cells are defined in figure 1. Under the null (scenario #0), the adaptive design controls the total Type 1 error below 2.5%, stops early for futility 63% of the time, and the average number of randomizations is 361. If the effect is uniform across cells—and therefore also over cumulated subgroups—(scenario #2), the fixed sample design is optimal, but the adaptive design results in only a small loss of power (from 89 to 80%). When the effect size distribution across the subgroups is in accord with the biological assumptions (scenarios #2 and 3), so that the effect in the cumulated subgroups declines as more cells with a null effect are added in, the adaptive design performs much better (higher power and smaller expected sample size) than the fixed sample, conventional trial. The adaptive design also performs well compared to a non-adaptive, fixed sample that includes efficient multiple comparisons-adjusted testing for effect in subgroups at the end of the study.²⁻ Note that in all of the tabulated examples (other than the first, which is the null case), the overall population null is false, because the effect is positive in some cell or cells, and the effect is null in the others, so that the effect never becomes negative. If we allow negative effects in any of the cells, then the advantages of the adaptive design over the conventional design grow even larger.

11. Exploratory analyses of the primary outcome after trial completion

11.1. DAWN eligible vs DAWN ineligible

The primary efficacy and safety analyses (mRS shift at 90 days, mRS 0-2 at 90 days, death rate and SICH rate) will be performed in “DAWN eligible” vs. “DAWN-ineligible” subgroups. The DAWN-ineligible subgroup will be defined as any patient who has one or more of the following characteristics:

- Pre-stroke mRS >1
- Baseline NIHSS <10
- Baseline ischemic core > 51 ml (based on Central Lab reading)
- Age >80 with baseline ischemic core > 21ml (based on Central Lab reading)
- Age < 80 and NIHSS < 20 and Baseline ischemic core >31 ml (based on Central Lab reading)

All other patients will be considered “DAWN eligible”.

11.2. Analysis in Subgroups Defined by Sex, Race, and Ethnicity

The primary efficacy analysis will be repeated in subgroups defined by sex, race (White, African-American, Asian, Other), and ethnicity (Hispanic, Non-Hispanic). The treatment effect and corresponding 95% confidence intervals will be estimated, as described in Section 12.2.

11.3. Analysis adjusting for covariates

Please note: if the final analysis is performed with 200 patients or less, we will still perform the analysis of covariates as described below, but acknowledge that the value of this analysis may be limited because of the small sample size.

An additional ~~The secondary~~ analysis of the primary endpoint will be the same rank-based analysis comparing the distribution of the 90-day Rankin scores between treatment groups while stratifying for prognostically important covariates. The Generalized Cochran-Mantel-Haenszel test (CMH) is the analogue to the WMW, designed to test against the alternative that there is a uniform shift of size “delta” in the Rankin score distribution from one group to the other after stratification. The CMH test will be stratified by age, baseline NIHSS Score, Baseline Ischemic Core Volume (based on Central Lab reading), and Time from symptom onset to randomization (using the same categories used for the stratified randomization). A rich body of clinical research has established the prognostic importance of these covariates. Cut-points will generally follow those used in the randomization, but may be altered depending on the numbers of patients enrolled into each category.

To assess the relationship between important covariates and the size of the treatment effect, a model based regression analysis will be performed. Specifically, we will create a multivariable ordinal regression model with the 90-day Rankin outcome as the dependent variable and use a literature based model-building process. The model building process will include preliminary variable selection, a final model selection, and a final model assessment. We will evaluate each candidate predictor for proportionality and linearity (model assumptions). The correlation between candidate continuous variables will be analyzed. Variables with a rho coefficient higher than 0.7 will not be jointly entered into the final model building process - only the variable with the lowest p-value if it is less than 0.1. The base model will include treatment group. During the preliminary variable selection, potential predictors will be added to the base model one at a time to obtain a p-value for each. Variables considered candidates for the final model will have a p-value ≤ 0.1 . At the final model selection, variables will be added sequentially starting with the variable with the lowest p-value from the group of candidate predictors. The criteria for keeping a variable in the final model will be a p-value ≤ 0.05 . Each time a predictor is kept in the final model all previously added variables will be re-assessed. Any previously entered variables whose p-value has increased above 0.05 will be dropped from the model. All qualifying variables from the preliminary selection phase will be considered along with their two-way interaction with treatment assignment ~~site of arterial occlusion~~. A statistical assessment of the model will be examined by the shrinkage statistic. A shrinkage statistic below 0.85 will indicate the model is overfitting the data, and the number of predictors should be reduced. The following variables will be considered for inclusion in the adjusted model:

- Age
- Baseline NIHSS score

- Baseline ischemic core volume
- Time from symptom onset to randomization
- Sex
- Admission SBP
- Baseline glucose

11.4. Analysis under the as-treated principle

Due to the nature of the study, some subjects in the endovascular treatment plus medical management group may not receive endovascular therapy. One instance when this may occur is if symptoms resolve spontaneously between randomization and start of the endovascular procedure. Another potential reason is the absence of an arterial occlusion on the baseline angiogram (due to spontaneous recanalization). Although it would be rare, we may also see some subjects who are randomized to the medical management only group, but are treated with endovascular therapy. Therefore, after completion of the study and if there is a crossover population, we will repeat the primary analysis (the Wilcoxon-Mann-Whitney measure of superiority) ~~within the adaptive design sample~~, but with patients categorized under the as-treated principle.

11.4.1. Group assignment under the “as-treated” principle

- **Endovascular therapy:** Patients who present to the endovascular suite (cath lab) and undergo a femoral puncture within 24 hours after time of onset of the qualifying stroke are assigned to the endovascular treatment arm under the “as-treated” principle.
- **Medical therapy:** Patients who do not meet the above criteria for endovascular treatment are assigned to the medical treatment arm under the “as-treated” principle.

12. Analyses of secondary efficacy outcomes

12.1. Analyses of secondary clinical efficacy outcome: functional independence

The effect of endovascular treatment will be assessed on a secondary clinical efficacy outcome: functional independence at 90 days, defined as an mRS score ≤ 2 at day 90. Results will be expressed as an unadjusted risk ratio with its 95% confidence interval and p-value.

12.2. Analyses of functional independence in subgroups

The unadjusted effect of endovascular treatment on the secondary efficacy outcome “functional independence” (mRS ≤ 2) will be analyzed in the following subgroups, assuming sufficient numbers of subjects are enrolled in each subgroup (minimum required is 10% of the total sample size in each subgroup). Results will be expressed as unadjusted risk ratios with their 95% confidence intervals and p-values. We will assess for differences in the treatment effect between subgroups (eg, age <70 vs ≥ 70) using a Breslow-Day test for binary covariates, or by including an interaction term into a logistic regression model for covariates with greater than 2 categories.

- Time from symptom onset to randomization (using the categories from the dynamic stratification: <9 , 9-12, >12 hr ~~endpoint from adaptive design~~)

- Baseline ischemic core lesion volume based on Central Lab reading (~~using cutpoints from adaptive design~~ using the three categories from the dynamic stratification: <10.0, 10.0-25.0, > 25.0 mL)
- Age at randomization (<70 and \geq 70 years old)
- Baseline NIHSSS using the categories from the dynamic randomization (<13, 13 - 18 and >18)
- Baseline ASPECTS based on Central Lab reading (<8 and \geq 8)
- Primary occlusion site based on Central Lab reading (M1 and ICA)
- IV rt-PA treatment (yes and no)
- Patient selection (CTP versus MRI)
- ~~Collateral grade (0-1, 2, 3)~~
- Method by which symptom onset time was determined (last known well vs exact time of symptom onset)
- Sex
- Race (White, Non-white)
- Ethnicity (Hispanic and Non-Hispanic)
- Baseline atrial fibrillation (yes and no)

12.3. Analyses of imaging efficacy outcomes

We hypothesize that endovascular treatment improves radiological outcomes. We will compare four imaging outcomes between treatment groups:

- The proportions of successful reperfusion, where successful reperfusion is defined as a >90% reduction in the volume of brain tissue with critical hypoperfusion ($T_{max} > 6$ sec after artifact removal by core lab) between baseline and 24 hours. Results will be expressed as an unadjusted risk ratio with its 95% confidence interval and p-value.
- The proportion of subjects with recanalization of the primary arterial occlusive lesion (ICA or M1) at the time of the 24-hour follow-up scan is compared between the two treatment arms. Results will be expressed as an unadjusted risk ratio with its 95% confidence interval and p-value.
- Infarct volumes at 24 hours, defined as the lesion volume as outlined on the 24-hour DWI (or CT if DWI not performed). The 24-hour endpoint is based on data demonstrating that assessment of infarct volume at 24 hours captures the effect of reperfusion therapies on infarct growth and predicts outcomes similarly to day 90 infarct volumes.^{8, 9} The Wilcoxon rank-sum will be used to compare the treatment groups.
- Ischemic lesion growth between baseline and 24 hours defined as the difference between the baseline ischemic core lesion volume and the 24-hour DWI lesion volume (or CT if DWI not performed). Absolute lesion growth will be calculated by subtracting the baseline ischemic core lesion volume from the 24-hour DWI lesion volume. DEFUSE 2 demonstrated a substantial reduction in infarct growth among Target mismatch patients treated in the 6-12 hour time-window who achieved early reperfusion: median growth 0.5 ml (IQR: -2 – 10) with reperfusion (n=23) vs. 39 ml (IQR: 18-121) without reperfusion (n=13), $p < 0.001$. The Wilcoxon rank-sum will be used to compare the treatment groups.

We also hypothesize that final infarct volumes (assessed on 24-hour DWI) can be predicted based on baseline ischemic core and critically hypoperfused tissue volumes. To analyze this relationship, subjects will be divided into two groups: 1) patients with reperfusion, defined as >90% reduction in the $T_{max} > 6$ -second lesion volume between baseline and 24 hours (or TIC1 2b–3 at end of procedure if perfusion imaging inadequate) and (2) a “no reperfusion” group, defined as <10% reperfusion at 24 hours. Spearman’s rho will be used

to assess correlations between: 1) baseline ischemic core volume and 24-hour DWI lesion volume in patients with reperfusion; and 2) the union of baseline ischemic core volume and 24-hour critically hypoperfused lesion volume (Tmax6) and 24-hour DWI lesion volume in patients without reperfusion.

12.4. Additional pre-planned analyses

- Assessment of the effect of endovascular therapy on the following endpoints will be assessed: (1) >8 point improvement on the NIHSS between baseline and 24 hrs or an NIHSS score of 0-1 at 24 hrs; (2) discharge destination (home vs other); (3) number of days spent at home during first 90 days after stroke; and (4) mRS score at day 30 (both with an ordinal analysis and dichotomous at 0-2).
- Assessment of the interaction between time-to-treatment randomization and the effect of endovascular therapy (expressed as the common odds ratio for shift on the 90-day mRS)
- Assessment of imaging predictors of growth of the ischemic lesion between baseline and 24 hours stratified by treatment allocation and reperfusion status. Specific baseline imaging predictors that will be evaluated include the ratio of the Tmax 10 / Tmax 6 lesion volumes and the CBV value within the Tmax 6 lesion volume.

13. Safety analyses

All safety outcome measures are analyzed under the as-treated principle. Under this principle, each subject is analyzed according to the treatment that the subject received. The external Medical Safety Monitor and the DSMB will monitor safety variables throughout the study at frequent intervals. Details of this process are specified below.

13.1. Monitoring of deaths and sICH

The primary safety endpoints are deaths and sICH rates; these will be monitored quarterly. Symptomatic ICH or death rates that exceed pre-specified thresholds will trigger a meeting of the DSMB. The DEFUSE 3 study will be placed on hold if it is determined with 95% probability that either:

- 1) the symptomatic ICH rate (NIHSS worsening of 4 or more points associated with ICH) within 36 hours of randomization exceeds 10% in the endovascular group; OR
- 2) the 90-day mortality rate exceeds 20% in the endovascular group.

The study will remain on hold until the investigators and the DSMB can conduct a review of events and make a determination on the continuation of the trial.

13.2. Summary of Adverse Events (AEs) and Serious Adverse Events (SAEs)

All AEs and SAEs are summarized by MedDRA preferred term (as coded based on the AE CRF) and by treatment group in terms of frequency of the event, number of subjects having the event, timing relative to randomization, and relatedness to the intervention.

For the following specific events, the proportions by treatment group, as well as the RR and corresponding 95% confidence interval, are provided:

- Stroke related mortality within 90 days of randomization

- sICH within 36 hours of randomization
- Significant neurological deterioration prior to discharge (defined as an increase of 4 points or more on the NIHSS)
- PH1 and PH 2 rates on the 24 hr scan

At the end of the study, the cumulative incidences of these events are compared between the two treatment groups using Pearson Chi-Square. Fisher's exact test will be used if any cells are 5 or less.

All brain scans obtained prior to discharge will also be assessed for PH by the core lab.

14. Coordination between Stanford and MUSC statistical teams

The primary statistician for the study is Dr. [REDACTED] at Stanford. He is a voting member of the DEFUSE 3 Executive Committee (EC). Dr. [REDACTED] will be blinded to all outcome data during the study. Dr. [REDACTED] ~~will become unblinded~~ Upon database lock of the clinical database, Dr. [REDACTED] will receive unblinded data and will conduct the final analyses. Drs. [REDACTED] and [REDACTED] will be responsible for developing and writing the statistical analysis plan (SAP) prior to the initiation of the study, and SAP amendments, if any, during the study. The statistical team at MUSC, led by Dr. [REDACTED] will be unblinded throughout the study. The MUSC team will ~~implement the adaptive design algorithm developed and written by Dr. [REDACTED]~~ conduct and independently validate the interim analyses according to the SAP; generate Open and Closed Reports for the DSMB and interact with the DSMB in closed sessions; and collaborate with Dr. [REDACTED] on validation of final analyses. After database lock, the MUSC statistical team will create the public use datasets (PUDS) and submit them to the NINDS.

15. References

1. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975;31:103-115
2. Lai TL, Lavori PW, Liao OY. Adaptive choice of patient subgroup for comparing two treatments. *Contemp Clin Trials*. 2014;39:191-200
3. Lai TL, Shih M-C. Power, sample size and adaptation considerations in the design of group sequential clinical trials. *Biometrika*. 2004;91:507-528
4. Howard G, Waller JL, Voeks JH, Howard VJ, Jauch EC, Lees KR, et al. A simple, assumption-free, and clinically interpretable approach for analysis of modified rankin outcomes. *Stroke*. 2012;43:664-669
5. Nogueira RG, Lutsep HL, Gupta R, Jovin TG, Albers GW, Walker GA, et al. Trevo versus merci retrievers for thrombectomy revascularisation of large vessel occlusions in acute ischaemic stroke (trevo 2): A randomised trial. *Lancet*. 2012;380:1231-1240
6. Saver JL, Jahan R, Levy EI, Jovin TG, Baxter B, Nogueira RG, et al. Solitaire flow restoration device versus the merci retriever in patients with acute ischaemic stroke (swift): A randomised, parallel-group, non-inferiority trial. *Lancet*. 2012;380:1241-1249
7. Pereira VM, Gralla J, Davalos A, Bonafe A, Castano C, Chapot R, et al. Prospective, multicenter, single-arm study of mechanical thrombectomy using solitaire flow restoration in acute ischemic stroke. *Stroke*. 2013;44:2802-2807
8. Campbell B, Purushotham A, Christensen S, Desmond P, Nagakane Y, Parsons M, et al. The infarct core is well represented by the acute diffusion lesion: Sustained reversal is infrequent. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*. 2012;32:50-56

9. Campbell BC, Tu HT, Christensen S, Desmond PM, Levi CR, Bladin CF, et al. Assessing response to stroke thrombolysis: Validation of 24-hour multimodal magnetic resonance imaging. *Arch Neurol*. 2012;69:46-50