STATISTICAL ANALYSIS PLAN (SAP)

BEnefits of **S**troke **T**reatment Delivered Using a **M**obile **S**troke **U**nit Compared to Standard Management by Emergency Medical Services: The **BEST-MSU** Study

Houston Data Coordinating Center

PI: Jose-Miguel Yamal, PhD

Co-Investigator: Suja S. Rajan, PhD

UTHealth, School of Public Health

Houston, TX 77030

June 2017

updated March 2018 (sensitivity analysis of utility weights, adjusted analysis for TM subgroup analysis)

updated January 2019 (include a pre-specified subgroup of EMS arrival <1 hr vs >1hr from last seen normal)

updated February 2020 (include clarification of the modified ITT analysis and sensitivity analyses)

updated April 2020 (included 30% improvement from baseline to 24 hr NIHSS as a secondary outcome)

updated June 2020 (clarified the primary analysis as a regression model, including site adjustment and alternative modeling if assumptions are not met; added new utility weights)

updated April 2021 (included propensity score analysis as a posthoc analysis for all outcomes; included analysis of all enrolled cohort and adjudicated enrolled excluding stroke mimics and hemorrhages)

Table of Contents

1. STUDY OVERVIEW

1.1. Objective and Study Design

The primary goal of this project is to carry out a trial comparing pre-hospital diagnosis and treatment of patients with stroke symptoms using a Mobile Stroke Unit (MSU) with subsequent transfer to a Comprehensive Stroke Center (CSC) Emergency Department (ED) for further management, to standard pre-hospital triage and transport by Emergency Medical Services (EMS) to a CSC ED for evaluation and treatment (Standard Management-SM).

There are many ways that use of a MSU might prove valuable in stroke patients, but we will focus on acute ischemic stroke (AIS) and treatment with IV tissue plasminogen activator (tPA) within 4.5 hours of symptom onset since that is the most evidence based effective emergency treatment for the most prevalent stroke diagnosis. We hypothesize that the MSU pathway will produce an overall shift towards earlier evaluation and treatment, particularly into the first hour after symptom onset, leading to substantially better outcome. We will also explore the hypothesis that as a result of improved clinical outcomes resulting from earlier treatment, the costs of a MSU program will be offset by a reduction in the costs of long term stroke care and increase in quality adjusted life years, thereby supporting more widespread use of this technology. To make MSU deployment more practical, we will confirm that a Vascular Neurologist (VN) on board the MSU can be replaced by a remote VN connected to the MSU by telemedicine (TM) thereby reducing manpower requirements and costs.

The successful completion of this project will provide data on important outcomes and costs associated with the use of MSU vs SM in the United States (U.S.) that will determine the value of integrating MSUs into the pre-hospital environment that would be more generalizable throughout the country. Therefore, the proposed study is the necessary step in a process that may dramatically modify the way that acute stroke patients are managed.

This is a prospective multicenter cohort study with randomized deployment weeks and blinded assessment of both trial entry and clinical outcomes.

2. DEFINITION OF TARGET POPULATION AND STUDY SAMPLES

2.1. Target Population

No. of Clinical Sites: 6
No. of subjects:

To be assessed for eligibility	(n = 4900)
To be enrolled	(n = 1845)
To be analyzed (“tPA eligible”)	(n = 1038)

Main criteria for inclusion:

1. Criteria for MSU team to **enroll** a patient into the study (to be determined pre-hospital on both MSU and SM weeks)

a. Last seen normal possibly within 4hr 30 min
b. History and physical/neurological examination consistent with acute stroke
c. No definite tPA exclusions per guidelines, prior to CT scan or baseline labs
d. Informed consent obtained from patient (if competent) or legal representative. Pre-hospital management and treatment, including IV tPA, will not be delayed for consent; however, consent in both MSU and SM patients must eventually be obtained for data to be retained for analysis.

2. Criteria for **tPA-eligibililty** (to be determined pre-hospital on MSU weeks, and after ED assessment on SM weeks, and confirmed by blinded adjudication)
   a. Meeting tPA inclusion and exclusion criteria per guidelines after CT scan, baseline labs, and clinical re-evaluation

2.2. Study Outcomes

2.2.1. Primary Outcomes

- The utility-weighted modified Rankin Scale (mRS) at 90 days, comparing patients found eligible for tPA (based on a blinded review of the patient's chart, regardless of whether they were treated or not) on MSU weeks compared to patients on SM weeks.

2.2.2. Secondary Outcomes

- Comparing patients found eligible for tPA (based on a blinded review of the patient's chart, regardless of whether they were treated or not) on MSU weeks compared to patients on SM weeks.
  - ordinal (shift) analysis of mRS at 90 days, and
  - proportion of patients achieving 90-day mRS 0,1 vs 2-6
  - 30% improvement from baseline to 24hr NIHSS

- The agreement between the VN on board the MSU with a VN remotely assessing a suspected stroke patient for treatment with tPA via TM in the MSU, and the rate of technical failures in conducting the TM consultation. N.B. Patients will include all enrolled patients on MSU weeks considered for tPA treatment.

- An exploratory cost-effectiveness analysis (CEA) of MSU versus SM using the Incremental Cost Effectiveness Ratio and Incremental Net Benefit estimate will be performed. N.B. The exploratory CEA will include all enrolled patients on MSU and SM weeks found eligible for tPA (based on a blinded review of the patient's chart, regardless of whether they were treated or not)

- Comparing all patients treated with tPA (whether or not adjudicated as tPA eligible) on MSU weeks compared to patients on SM weeks.
    - Utility-weighted modified Rankin Scale (mRS) at 90 days
    - ordinal (shift) analysis of mRS at 90 days, and
    - proportion of patients achieving 90-day mRS 0,1 vs 2-6
    - 30% improvement from baseline to 24hr NIHSS

- Comparing enrolled patients treated with tPA within 60 minutes of LSN onset according to published guidelines on either MSU or SM weeks, compared to similar patients treated 61-270 minutes after onset, adjusting for any imbalances in stroke severity (baseline NIHSS) between the groups at the time of treatment. N.B. Patients will include only those patients actually treated with tPA based on the final determination of the time LSN, and will include only patients meeting all inclusion and exclusion criteria.
    - utility-weighted mRS at 90 days,
    - ordinal (shift) analysis of mRS at 90 days
    - proportion of patients achieving 90-day mRS 0,1 vs 2-6
    - 30% improvement from baseline to 24hr NIHSS
    - Instead of dichotomizing into two groups based on time from LSN to tPA, logistic regression of 90-day mRS 0,1 vs 2-6, using a restricted cubic spline for time from onset to treatment, with visualization of spline term compared with the odds ratio

- Comparing all patients treated with IAT (separate analyses for those adjudicated as tPA eligible, all tPA treated, or all IAT with or without tPA) on MSU weeks compared to patients on SM weeks.
    - utility-weighted mRS at 90 days
    - ordinal (shift) analysis of mRS at 90 days
    - proportion of patients achieving 90-day mRS 0,1 vs 2-6
    - 30% improvement from baseline to 24hr NIHSS

- The time from LSN to tPA treatment on all patients treated within 4.5 hours of LSN on MSU weeks compared to similarly eligible patients on SM weeks. N.B. Patients will include all enrolled patients actually treated with tPA (or on SM weeks, eligible for tPA treatment) meeting all inclusion and exclusion criteria, and based on the final determination of time of LSN. One analysis will compare the median times. A second analysis will also capture the patients who were eligible but did not receive tPA because it was too late, categorizing time into the following groups (e.g., 0-60min, 61-90min, 91min-180min, 181-270min, eligible but no tmt because>270).

- Of the enrolled patients that were eligible for treatment with tPA (according to

published guidelines) on MSU weeks compared to SM weeks, the percent that were treated within 4.5 hours and within 60 minutes of LSN.

- The time from LSN, from alarm time, and from ED arrival to start of endovascular procedure (intra-arterial thrombectomy-IAT) in patients who meet pre-specified criteria for IAT on MSU weeks compared to SM weeks. N.B. All patients receiving IAT will be included in this outcome.

- The proportion of all tPA-eligible patient having IAT on MSU weeks compared to SM weeks

- The median/mean time from LSN to tPA therapy decision on all patients considered for treatment within 4.5 hours of LSN on MSU weeks compared to SM weeks. N.B. Patients will include all enrolled patients meeting inclusion criteria whether or not treated with tPA.

- Time between 911 call and onset of etiology-specific BP management on MSU weeks compared to SM weeks. N.B. Patients will include all enrolled patients.

2.2.3. Safety Outcomes

- The incidence of symptomatic intracranial hemorrhage (sICH) in enrolled tPA treated patients on MSU weeks compared to SM weeks (Symptomatic intracranial hemorrhage defined as any intracranial blood accumulation associated with a clinical deterioration of $\geq$4 points of the NIHSS for which the hemorrhage has been identified as the dominating cause of the neurologic deterioration) N.B. Patients will include all patients treated with tPA, whether or not they meet all inclusion and exclusion criteria.

- Mortality. N.B. All enrolled patients signing informed consent will be included in this endpoint and followed until 1 year.

- The incidence of stroke mimics and transient ischemic attacks (TIAs) in tPA-treated patients, and also in tPA-eligible patients, on MSU weeks compared to SM weeks. N.B. SM patients deemed eligible for tPA on their pre-hospital assessment who then completely recover by the time of arrival in the ED will equal the excess incidence of TIAs treated on the MSU pathway.

3. GENERAL STATISTICAL CONSIDERATIONS

3.1. Randomization and Analytic Cohorts (The process is described in detail in the protocol)

Weeks when the MSU is available or not are randomly selected. Stroke events are orthogonal to whether the MSU was being deployed or not that week and thus participants will be randomly

entered into either the MSU or SM groups depending on when their stroke occurs.

The primary analytic cohort is based on a modified intention-to-treat (ITT) analysis where the subject will be assigned to the group that they were enrolled in (e.g. if a patient was enrolled using SM, they would be assigned to the SM group) and adjudicated (by the blinded adjudicator) to be tPA eligible. The usual ITT includes every subject who is randomized according to randomized treatment assignment. In this study, all patients within each group who are adjudicated as tPA eligible by an adjudicator blinded to group assignment are included. The randomized assignment is not conducted for each patient, rather we generally alternated weeks to be either MSU or SM weeks, which is independent of when a subject randomly has a stroke and calls 911. Therefore, this may be considered a cluster-randomized trial where the cluster is the days when the MSU is available and the other cluster is when MSU is not available. There is not anything clinically important to set the cluster of when the MSU was available or not as a week (e.g., an alternative design could set one week as having MWF as MSU days and TTH as SM days and the next week as the opposite), but this made it convenient to set work schedules and to have a similar amount of time dedicated to recruitment of MSU and SM subjects and there is not a scientific nor statistical rationale suggesting that the clusters would be related to the patient's outcomes and intervention effect. Patients are, in a sense, "randomly" allocated into the clusters based entirely on when they happen to have their stroke in relation to the prospectively determined cluster allotment of whether the MSU is available or not. Furthermore, in order to optimize the utilization of the MSU, some cities have 2 sites enrolling patients at the same time, with one site running the MSU and other enrolling SM patients and then they switch the next week.

There are a few cases when the MSU was not available during an "MSU week" (e.g. the unit is out of service on another call, had to be serviced for an oil change, staff were sick and therefore unable to come in) and stroke patients that were treated using standard management were enrolled into the study by the study team into the SM arm. These few subjects will be included in the primary analysis in the SM arm, but moved to the MSU arm in a sensitivity analysis (see section 5.1.3). The decision to include them in the primary analysis is based on a November, 2019 comparison of the SM subjects who were enrolled during an "MSU week" compared to the SM subjects enrolled during an "SM week". Baseline characteristics (age, sex, ethnicity, race, pre-stroke mRS, baseline NIHSS, tPA treatment, time from LSN to tPA bolus, endovascular treatment, and DTGP) were similar between the groups, confirming our belief that there should not be any added bias for including them in the primary analysis. The benefit of including them is to improve the MSU:SM ratio and to increase the chance of recruiting subjects according to the projected timeline. However, this analysis will be repeated at the end of the study to confirm that no significant differences exist between these two SM populations before including them in the MSU arm.

In response to peer review, post hoc analyses were added to look at two additional cohorts: (1) all enrolled, regardless of adjudication; (2) all adjudicated enrolled, excluding hemorrhages.

3.2. Blinding

Blinded assessment of both trial entry, tPA-eligibility, and study outcomes. All patients are screened for trial enrollment during their pre-hospital evaluation and management by the same investigators

on both MSU and SM weeks to ensure that comparisons are made between similar patients, using similar criteria, at a similar stage of illness. For enrolled patients, criteria for study enrollment and tPA treatment are subsequently reviewed by a vascular neurologist (VN) blinded to MSU vs SM assignment and not otherwise involved in study management or analysis. The blinded VN determines from a dedicated "adjudication form", omitting any time data or other information that would produce unblinding, if the patient meets criteria for study enrollment and for tPA treatment. For comparing outcomes between MSU and SM, we will only include tPA-eligible patients on both MSU and SM weeks, whether or not actually treated, based on this blinded review. Investigators obtaining all outcomes are blinded to treatment allocation.

3.3. Multiplicity

No adjustments for multiple comparisons will be made. However, the secondary analyses will be interpreted with caution.

4. SAMPLE SIZE DETERMINATION

4.1. Sample Size for the Phase III trial

The power of this trial was based on the difference in primary outcome, 90 d uw-mRS. Based on preliminary data, we expected 1.8 times as many MSU as SM patients because when we began the study, on SM weeks some patients were occasionally taken by EMS to non-participating stroke centers where they could not be enrolled into the study. On MSU weeks, these patients would be transported in the MSU only to participating hospitals and therefore enrolled. Subsequently, we have incorporated these non-participating hospitals into the study, thereby mitigating this gap and the groups are now balanced. With a sample size of 693 total tPA-eligible patients (446 MSU and 247 SM patients, assuming 10% lost to follow-up [LTF]), the study will have 80% power with a 0.05 Type I error to detect a difference between groups of 0.09 in the mean uw-mRS using a two-sample t-test. This difference is plausible and important. In a re-analysis of 11 acute stroke studies8, the difference in mean 90d uw-mRS between groups ranged from 0.024-0.25, with most positive trials in the range of 0.1. In the NINDS tPA trial, 90d uw-mRS difference was 0.09 between tPA and placebo.

In March, 2018, Dr. Grotta, blinded to study data, requested, and PCORI approved, an increased sample size to 1095 patients from the 693 initially requested, and to allow three additional sites to be added. This request was based on our reassessment of anticipated difference in 90 day uw-mRS based on a.) results of the Berlin non-randomized study which showed a 0.07 difference between MSU and control patients, b.) results of the DAWN trial which was the first completed study to use the uw-mRS, and c.) reanalysis of a substantial number of completed stroke trials where conventional mRS outcomes were translated to uw-mRS (see figure below). In that analysis, Broderick et al found that the smallest clinically meaningful difference was 0.04[1]. We based our initial sample size of 693 tPA eligible patients on the ability to detect a 0.09 point difference which was the same as between tPA and placebo in NINDS. The endovascular studies found a >0.10 point difference. Based on these pieces of information which were not available when we designed our study, Dr. Grotta reassessed

the anticipated difference between groups if the MSU produces a substantial reduction in time to treatment, and felt that a difference of 0.07 is a more realistic goal. Dr. Yamal did not participate in that decision since he is unblinded.

Assuming a 3:2 (1.5) imbalance, 5% LTFU, and using the pooled standard deviation of STEMO & No STEMO group (sd=0.385), numbers of patients needed to detect a difference of 0.07using a 2-sample t-test is N=1038. Our LTFU so far has been around 5% so we expect this assumption to be reasonable. PCORI has agreed to the increase in sample size and sites sufficient to detect a 0.07 difference.

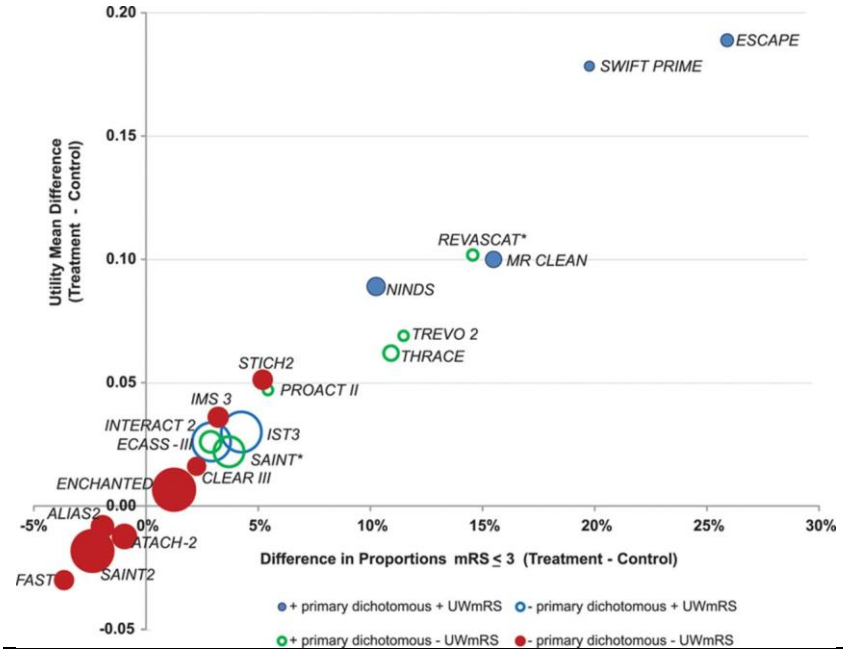

Figure. Reanalysis of a substantial number of completed stroke trials where conventional mRS outcomes were translated to uw-mRS. Effect sizes reported. From Broderick, et al.

4.2. **Sample Size Estimation for Cost Effectiveness Analysis**

We will perform an exploratory cost analysis using the cost data collected during this study. Based on the sample size estimation outlined in Willan et al[2], and cost and QALY estimations from past studies[3-6], we estimated a range of sample sizes that will be required for a formal CEA. The lowest and highest observed change in QALY in the literature was 5-20%; similarly observed change in cost was 10-25%. Based on these the sample size requirement in the most optimistic case was 96 patients (48 in each group) and in the most conservative case was 740 patients (370 in each group) for a power of 80% and p-value if 0.05. Approximately 50% of the patients for whom the MSU is dispatched, and who meet inclusion criteria for enrollment into the study, will receive tPA. Hence, the total number of patients used for the CEA will have to be between 192 and 1480 patients. Even though the current study probably will not meet the

sample size requirement for the conservative case, it will help establish the expected cost and QALY changes for the MSU intervention (which have never been estimated before).

5. ANALYSIS PLAN

5.1. Phase III Trial Analysis

5.1.1. Treatment Group Comparability at Baseline

Although the random enrollment of participants to the two treatment arms and blinded review of tPA eligibility should ensure comparability with respect to known and unknown variables, imbalance may occur by chance. Descriptive statistics for baseline characteristics known or suspected to be associated with outcomes will be prepared for the two treatment groups for all randomized as well as all deemed "eligible for tPA" based on the blinded review. Chi-square statistics and Wilcoxon rank sum tests will be used to evaluate baseline differences between the arms for categorical and continuous variables, respectively. Any variables with baseline differences will be included in secondary adjusted analyses. Also, completers will be compared to non-completers (loss to follow-up for 90 mRS) on these baseline variables to indicate whether missingness may be considered random.

5.1.2. Primary Clinical Analysis

The mean uw-mRS at 90d along with corresponding two-sided 95% confidence intervals will be compared between groups using a two-sample t-test or Wilcoxon rank sum test if the assumption of normality does not hold. Although the mRS is an ordinal outcome, the difference between the uw-mRS categories has clinical significance and the t-test assumption and central limit theorem are likely satisfied. The primary analysis of uw-mRS will be adjusted for baseline uw-mRS, site, any baseline covariates that are different between the two groups, and covariates associated with mRS, including baseline NIHSS, age, pre-morbid mRS, and previous TIA/stroke, in a linear regression model. If the assumptions of the model are not satisfied, a restricted cubic spline will be used to model baseline continuous variables (NIHSS and age). If the linear regression with splines does not fit well, we will use ordinal logistic regression to adjust for the variables. If the proportional odds assumption fails, we will use logistic regression with mRS 0-1 vs 2-6 as the primary analysis. Assuming that the primary analysis doesn't use the following models, sensitivity analyses of the primary outcome will be conducted including ordinal (shift) analysis using a proportional odds model and proportion achieving a dichotomized outcome of mRS 0-1 vs 2-6 using binary logistic regression.

In response to peer review, we added post-hoc propensity score analyses using propensity scores as an alternative way to reduce any effects of confounding to estimate the effect of MSU group on dichotomized mRS (0-1 versus 2-6). The individual propensities for enrolling into the MSU versus EMS groups were estimated using a separate selection multivariable logistic regression model with variables site, baseline NIHSS, pre-stroke

mRS, age, Black race, gender, and dichotomized time from LSN to EMS/MSU arrival (>1hr versus ≤1hr). Standardized mean differences were used to assess covariate balance before and after weighting (all standardized mean differences were < 0.1). The predicted probabilities were used to calculate stabilized inverse probability weights (IPW). The 90-day uw-mRS was further described using means and standard deviations and by fitting a univariate linear regression with outcome (90-day uw-mRS), covariate MSU group, and IPW according to the propensity score. IPW analyses of all enrolled patients was added post-hoc to assess the chances of post-enrollment selection bias and to align analysis with overall MSU vs EMS management with outcomes discharge mRS and 24hr NIHSS.

Utility weights

The sample size was originally designed using the Dawn trial utility weights that were derived based on a United Kingdom sample and using the 3-level version of EQ5D. EQ5D-5L has been in use for more than a decade. However, the corresponding population-level utility weights for 5L had not been developed for many countries, and most countries only have population-level weights for the much older EQ5D-3L. In 2019 a study was published by Pickard et al. conducted a survey to develop utilities based on a US population and using the 5-level version of EQ5D (EQ5D-5L), which is more relevant to the participants in the BEST-MSU study. Using their Probit model estimated parameters to calculate utilities, in June 2020 we fit a linear regression model with these utilities (using 90-day EQ5D-5L) as the outcome and the 90-day mRS indicator variables as the independent variables to estimate our specific utility-weighted mRS. We also applied both the Dawn and the newly derived utility weights to the B-PROUD data and observed that results were consistent in the comparison of their mobile stroke unit data and their non-mobile stroke unit groups between these two weight choices. These were presented during the June 2020 study monitoring committee and were approved to use as the primary outcome of our trial. The weights based on the June 2020 data are presented in Table.

Table. Comparison of utility weights.

mRS	Dawn weights (previous MSU weights)	ENCHANTED trial weights	New proposed MSU
0	1	0.977	1
1	0.91	0.885	0.91
2	0.76	0.748	0.72
3	0.65	0.576	0.65
4	0.33	0.194	0.18
5	0	-0.174	0.05
6	0	0	0

5.1.3. Sensitivity Analysis of Primary Outcome

We will conduct sensitivity analyses using Dawn utility weights of the 90 day mRS. An additional sensitivity analysis will add an indicator of whether the SM subjects that were enrolled during an MSU week affect the treatment effect in a regression model by adding an indicator for these subjects. Also, in a sensitivity analysis, we will move these subjects into the MSU arm to check for consistency of results.

A further sensitivity analysis will remove subjects that were enrolled during the COVID-19 pandemic months (beginning of March, 2020).

5.1.4. Analyses of Ancillary Clinical Outcomes

We will also compare mRS at 90d (uw-mRS, Δ uw-mRS from baseline, ordinal (shift) analysis, and proportion achieving 0,1) in tPA treated patients treated within 60 minutes of LSN to patients treated 61-270 minutes, regardless of whether they were on MSU weeks vs. SM weeks. Patients on MSU weeks vs SM weeks will also be compared for differences in (a) the time from LSN to tPA treatment, (b) time from LSN, alarm time, and ED arrival to start of IAT, and for safety outcomes (i) mortality, (ii) symptomatic intracerebral hemorrhage, and (iii) incidence of tPA treated stroke mimics and transient ischemic attacks.

A logistic regression model will be used to compare 90 day mRS 0,1 vs 2-6 of patients treated with tPA within 60 minutes of symptom onset to similar patients treated 61-270 minutes after onset, adjusting for any imbalances in stroke severity (baseline NIHSS, age, premorbid mRS, and previous stroke/TIA incidence) between the groups at the time of treatment[8]. If baseline characteristics are significantly different between the two non-randomized groups, we will use propensity score analysis to limit potential bias. Also, we expect a higher incidence of spontaneous recovery (TIA) and stroke mimics may occur with earlier observation in the 0-60 minute group compared to those seen 61-270 minutes. The "natural history" of the incidence of spontaneous recovery and stroke mimics will be estimated from patients enrolled into the SM group, and will be considered in analyzing the comparison between patients treated with tPA within 0-60 min vs 61-270 min. Time to treatment and to endovascular procedures will be analyzed using Cox proportional hazards models, similarly to survival. Categorical outcomes will be analyzed using Fisher's exact test.

Unless there is sufficient power (predetermined before the analysis is begun) the approach to ancillary analysis will generally be the calculation of confidence limits on intervention group differences rather than formal tests of significance as the trial may not have high power to detect difference in all of these outcomes. However, these comparisons will add to the knowledge of the benefits and risks of the intervention.

5.1.5. Subgroup Analysis

Tests of effects within subgroups will be driven by clinical rationale. To reduce the potential for spurious results, we would test for a sub-group treatment interaction at a 0.2 critical level. Any subgroup analyses that are not pre-specified would be considered post hoc and reported as requiring confirmation in future studies. Estimates of the MSU effect will be obtained separately for pre-specified subgroups with significant treatment-by-subgroup interactions, using the methods described above. Pre-specified subgroups include (1) patients treated via TM versus on-site VN, (2) patients treated at various sites, (3) patients that had the EMS arrive (for SM) or MSU arrive (for MSU) within $\leq$1 hr and those that arrived >1hr of LSN, and (4) race. For (3), time will also be considered as a continuous variable and the interaction between time and MSU/SM will be assessed with transformations or restricted cubic splines of time used if appropriate).

When doing the TM subgroup analysis, we anticipate that there may be demographic differences between sites that are doing TM versus onboard VN. For this analysis we will conduct regression models, adjusting for baseline NIHSS, age, pre-morbid mRS, time since last seen normal, and previous TIA/stroke, in a linear regression model.

Analyses of post-randomization sub-groups are subject to many biases. Thus any analyses of post-randomization sub-groups, such as those treated with IAT, would be considered on a case by case basis requiring tailored use of advanced statistical methods[9] and careful interpretation.

5.1.6. Missing Data

We expect no missing data for baseline measures. For 90-day assessments, extensive efforts will be made to ascertain the modified Rankin scores and mortality status, though we anticipate a 5% rate of lost to follow-up. We will perform several approaches for handling missing data. Characteristics of patients who are lost to follow-up will be compared to those that remain in the study to assess the degree of any selection bias, and sensitivity analyses will be performed to evaluate robustness of conclusions to the different missing data approaches. We will use multiple imputation for the final values assuming missing at random, depending on if any significant baseline differences exist between those observations that have a missing value or not. As sensitivity analyses we will report the data with and without imputation. Data will also be stratified according to their missing pattern (e.g., early termination, late termination, and follow-up completers) and variables representing these groups will be used as model covariates in adjusted analyses.

5.2. Cost Effectiveness Assessment

5.2.1. Approach and Methods used in Cost Analysis.

In order to establish an economic basis for a higher reimbursement from the healthcare

payers for dispatching an MSU the following aspects have to be established:

- Does the MSU improve the post-discharge stroke severity and consequently improve average patient QALYs? Higher cost for an intervention can be better justified if associated with improved patient outcomes.
- Does the MSU reduce post-stroke healthcare utilization and consequently costs for the healthcare payers? Reduction of post-stroke healthcare utilization will subsequently save costs for the healthcare payers who pay for these utilizations. By identifying whether the healthcare payers save costs for stroke management due to the use of MSU (and determining the amount of post-stroke cost savings) the study can provide scientific evidence for supporting additional Medicare reimbursements for an MSU dispatch.
- What is the magnitude of the incremental fixed costs associated with MSU and the per-patient incremental fixed cost due the ambulance outfitting, CT, other equipment, and telemedicine technology, staffing requirements and paramedic training? Establishing the magnitude of incremental fixed cost per patient will help determine the justifiable amount of increased reimbursements to agencies operating the MSU and providers supporting its telemedicine capabilities.

5.2.2. Sample used for Cost Analysis

The cost-effectiveness analysis (CEA) will include all enrolled patients on MSU and SM weeks who meet criteria for tPA treatment whether or not they are eventually treated with tPA. We estimate that approximately 50% of enrolled patients will receive tPA in the MSU and SM group. The non-tPA treated patients will probably not benefit much from MSU management and since the primary goal of the MSU is to ensure quicker administration of tPA, only those patients who meet criteria to receive tPA will be included in the cost analysis (for one year cost and QALY follow-up). The cost of operating the MSU for the remaining 50% of the patients who are not eligible for tPA administration will be included as fixed costs of operating the MSU, but these patients will not be followed-up once they are deemed ineligible to receive tPA inside the MSU or at the ED.

5.2.3. Perspective of the cost-effectiveness analysis (CEA)

The CEA will be performed from the perspective of the healthcare payers. If dispatching an MSU improves patient outcomes it should theoretically reduce post-stroke healthcare utilization and hence the reimbursement costs for the healthcare payers under the current payment policies, which do not include additional reimbursement for an MSU dispatch. If the study demonstrates improved effectiveness along with cost-savings or demonstrates improved effectiveness with limited increase in costs for the healthcare payers it will help justify the additional reimbursements for dispatching an MSU. This justification is vital for the financial viability of this high cost intervention and hence critical for the study.

5.2.4. Measure of Effectiveness

Stroke results in severe morbidity, disability and mortality in the American population.23 More than 70% of the stroke patients are unable to return to their pre-stroke life style,

activities of daily living and employment. Thus, stroke has a permanent impact on the patient's QOL, thereby necessitating the use of a patient-centered effectiveness measure that considers both the quality and quantity of a patient's life, and is not limited to physician reported clinical measures or survival. Hence, QALYs will be used as the effectiveness measure. QALYs will be obtained through utility-weight conversions using the EuroQol's EQ-5D measure. ED-5D is preferred due to its standardized ease of conversion to QALYs.33,38 We considered the use of other QOL measures like Neuro-QoL. After communication with the Neuro-QoL research team it was established that Neuro-QoL has not been validated for conversion to QALYs. In addition, Neuro-QoL involves the reporting of 18 adult domains in the form of separate T-scores which should not be combined to form a single QOL measure further limiting the feasibility of QALY conversion. Since costs analysis requires QALYs and not QOL measures, Neuro-QoL and similar stroke-specific QOL measures, which cannot be converted to QALYs, are not used in this study.

5.2.5. Measure of Cost

The cost components include: 1) The incremental fixed costs associated with the MSU 2) The index hospitalization costs 3) The post-discharge cost during the first year after the stroke episode 4) Life-time costs after the first-year. The incremental fixed cost (component 1) for the MSU group will include cost of additional outfitting required to convert an ambulance into an MSU, cost of additional staffing changes for the agency operating the MSU, provider/hospital-level infrastructure changes to accommodate the MSU, clinical staff training, EMS and dispatch training, and all trips performed by the MSU (whether they involve tPA eligible patients or not). The variable cost (cost per patient) will include components 2 to 4, and will be measured for all patients in the MSU and SM group who meet criteria for tPA treatment whether or not they are eventually treated with tPA. Microcosting (resources * local market value) will be applied to the estimation of incremental fixed cost (component 1) whereas gross costing (utilization * Medicare payments) will be used for the variable costs of post-stroke healthcare utilization in the first year (components 2 and 3). Life-time costs after the first year (component 4) will be simulated using Markov modeling based on evidence from the literature[10,11]. The fixed cost of CT scanners and telemedicine equipment will be amortized over the 10 year expected life of the equipment. Medicare reimbursement amounts for patients from different geographic areas will be adjusted to make them nationally representative by using the CMS geographic adjustment factor (for part A claims) and CMS geographic practice cost index (for part B claims).

5.2.6. Funding and Cost Analyses

The cost analyses will not be supported by the PCORI funding.

6. MONITORING FOR EFFECTIVENESS AND SAFETY

6.1. Overview

Interim analyses for safety (symptomatic hemorrhage), efficacy/futility (dichotomized mRS 0-1 vs. 2-6), and process (time from alarm until treatment decision) will be conducted when the 90-day mRS has been collected on 50% of the total number of patients that are adjudicated to be tPA-eligible.

6.2. Interim Analyses for Effectiveness

The efficacy interim analysis of the 90 day dichotomized mRS will be a 2-sample, 2-sided test of proportions using a Haybittle-Peto boundary (p=0.001). This will be conducted on the subset that are tPA-eligible based on the blinded adjudication.

6.3. Interim Analyses for Futility

The futility analysis of the 90 day dichotomized mRS (0-1 vs 2-6) will be a 2-sample, 1-sided, test of proportions. The futility analysis will compare patients in MSU weeks vs SM weeks (alpha=0.15). If we reject the null hypothesis that the percentage of favorable outcomes (mRS<2) in patients in the MSU weeks is greater than or equal to the percentage of favorable outcomes in patients in the SM weeks plus 10%, we conclude that completing the trial would likely be futile. The futility hypotheses are: $H_0: p_{MSU} - p_{SM} \geq \Delta$ versus $H_A: p_{MSU} - p_{SM} < \Delta$ where p_{MSU} and p_{SM} are the proportions of participants expected to have a favorable mRS outcome in the MSU and SM groups, respectively, and Δ denotes the 10% increase in favorable outcomes over SM considered clinically meaningful. This will be conducted on the subset that are tPA-eligible based on the blinded adjudication.

6.4. Safety Analyses

Rates of symptomatic hemorrhage will be compared using a Fisher's exact test (alpha=0.05). This will be conducted on all enrolled tPA-treated patients, excluding any that had an ICH on their baseline CT scan.

6.5. Process Analysis

Time from alarm to treatment decision will be compared using a one-sided Wilcoxon rank sum test (alpha=0.05) to test if the time is longer for the MSU arm. This will be conducted on the subset that are tPA-eligible based on the blinded adjudication. MSU-by-site interaction terms will be included in a regression model to test if these differ by site and if the interactions are

significant then within-site tests will be conducted.

7. REPORTING PROCEDURES

7.1. CONSORT Diagram

We will account for every subject randomized into the study using a CONSORT diagram.

7.2. Primary Reporting for the BEST-MSU Study

We will account for every subject randomized into the study using a CONSORT diagram. Primary reporting for the BEST-MSU study will follow the classic CONSORT Checklist items (see appendix).

7.3. SMC Reports

Standard format for SMC reports will be developed and sent to the SMC for review before the initial safety analyses are presented, and the format will be added as an appendix to this report.

7.4. Publications

Before the BEST-MSU CCC begins an analysis for a manuscript or presentation, the first author or writing group will have their hypotheses and analysis plan reviewed and approved by a designated team at the BEST-MSU DCC.

8. REFERENCES

1. Broderick JP, Adeoye O, Elm J. Evolution of the modified rankin scale and its use in future stroke trials. *Stroke*. 2017;48(7):2007-2012.

2. Willan AR. Sample size determination for cost-effectiveness trials. *Pharmacoeconomics*. 2011;29(11):933-949.

3. Mauldin PD, Simpson KN, Palesch YY, et al. Design of the economic evaluation for the interventional management of stroke (III) trial. *International Journal of Stroke*. 2008;3(2):138-144.

4. Ovbiagele B, Goldstein LB, Higashida RT, et al. Forecasting the future of stroke in the united states: A policy statement from the american heart association and american stroke association. *Stroke*. 2013;44(8):2361-2375.

5. Trogdon JG, Finkelstein EA, Nwaise IA, Tangka FK, Orenstein D. The economic burden of chronic cardiovascular disease for major insurers. *Health Promot Pract*. 2007;8(3):234-242.

6. Grotta JC, Albers GW, Broderick JP, et al. *Stroke: Pathophysiology, diagnosis, and management.* Elsevier Health Sciences; 2015.

7. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708-715.

8. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Wiley series in probability and statistics. *Applied Logistic Regression, Third Edition*. 2013:501-510.

9. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266(1):93-98.

10. Fagan SC, Morgenstern LB, Petitta A, et al. Cost-effectiveness of tissue plasminogen activator for acute ischemic stroke. NINDS rt-PA stroke study group. *Neurology*. 1998;50(4):883-890.

11. Tan Tanny SP, Busija L, Liew D, Teo S, Davis SM, Yan B. Cost-effectiveness of thrombolysis within 4.5 hours of acute ischemic stroke: Experience from australian stroke center. *Stroke*. 2013;44(8):2269-2274.

Appendix A: CONSORT Checklist

CONSORT CHECKLIST

Table. CONSORT 2010 Checklist of Information to Include When Reporting a Randomized Trial[a]

Section and Topic	Item No.	Checklist Item	Reported on Page No.
Title and abstract			
	1a	Identification as a randomized trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomization Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomization; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome	
	13b	For each group, losses and exclusions after randomization, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analyzed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
Comment			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalizability	21	Generalizability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

[a]We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomized trials, noninferiority and equivalence trials, nonpharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up-to-date references relevant to this checklist, see http://www.consort-statement.org.

http://jama.ama-assn.org/site/misc/auinst_chk.pdf