Statistical Analysis Plan

Sample Size Calculations: Our primary aim is to evaluate the impact of the mobility program on mobility in the year after hospital discharge. In our previous small VA-funded study of a mobility intervention, patients in the MP group had a mean LSA score that was approximately 10 points higher than in the UC group at four weeks post-discharge. It is our expectation that the current study will replicate these results; specifically that the 10 point difference will be apparent at 1-month post-discharge and persist thereafter. As described below, the primary statistical technique to address the primary aim will be mixed effects models. Assuming an exchangeable covariance structure across the repeated observations, a sample size of 190 participants provides greater than 90% power to detect a persistent 10-point difference in LSA scores at the $\alpha = .05$ level. We must also consider the reduced effective sample size due to clustering. It is a common practice to calculate the intraclass correlation (ICC) among clusters as a measure of how much more similar members of the same cluster are to each other than to the overall population. In this setting, we expect much more variability to come from individual patient factors than from the units. Using the data from our recently completed VA longitudinal study, we calculated an ICC for the LSA that rounded to 0. Assuming a larger ICC of 0.001 and subjects spread across 5 units, a sample size of 198 yields an effective sample size of 190 observations. In our recently completed VA longitudinal study during which patients were followed for 6 months after hospital discharge, 80% of the veterans completed at least 5 of the 6 monthly post-hospital assessments. We had 23 deaths and 3 withdrawals over the 6 months of follow-up among the 200 veterans in the study. Thus, in order to have an adequate (i.e.198) sample size at 12-months, 226 patients will be enrolled allowing for patient losses due to death or drop outs.

Data Management: Data collected from in-hospital assessments will be recorded by the research assistant on paper forms. Forms will be reviewed daily by the project coordinator for errors and completeness relating to the patient identification number and name, legibility of handwriting, appropriate form used, and missing data. The forms will then be entered into an Access database which will be programmed to include multiple error checking techniques for inconsistent data and out of range coding to ensure accuracy of the data entry.

Analytic Approach (for each specific aim):

General: Descriptive statistics including mean, standard deviations, distribution, maximum and minimum values for the study variables will be calculated. Data will be examined graphically and by summary statistics for outliers and distributional properties (e.g., normality, skewness) that may lead to transformations or nonparametric analyses where indicated.

To analyze differences in LSA while accounting for changes in units over time, we will use a stepped-wedge design (Figure 3, above), a variation of cluster randomization in which each cluster is eventually assigned to intervention.[58] By beginning with a pre-intervention period and implementing the interventions in 'steps' at set times, each unit serves as its own control, with pre-intervention observations used to estimate any underlying long-term or seasonal trends with precision so that effects from the intervention can be distinguished from these 'secular' trends. While the standard stepped-wedge design treats time in discrete blocks (the 'steps'), this trial will

be able to provide more robust estimates of secular trends because observations are linked to the date of each patient's admission, providing a much finer temporal resolution. To understand the secular trends, we will start by examining smoothed plots of LSA over time overall and stratified by unit to identify especially large temporal fluctuations that may indicate unanticipated changes in processes or personnel. We will then model LSA during the time prior to the initiation of the intervention using mixed models,[59] to account and test for heterogeneity among hospital units (in both means and trends) and possible nonlinearities. While secular trends must be accounted for in the final models, results from this preliminary analysis will inform whether potential trends can be considered to be linear, and whether all units can be considered to have the same trends over time.

Informed by the modeling of secular trends, the final models will add in indicator terms for whether the observation occurred before or after the intervention, in addition to the appropriate modeling of secular trends. Additional more sophisticated modeling will evaluate the trends in the outcomes over time in the post-intervention periods to see whether temporal trends differ pre-and post- intervention. All statistical tests will be two-sided and will be considered to be significant if the associated p-value is <0.05. All analyses will be intention-to-treat based on the unit's assignment at the time of a patient's admission, and missing data will be imputed using a multiple imputation strategy, based on the characteristics of the data.

AIM 1: To test the effectiveness of a mobility program on recovery to pre-hospital mobility status or better and reduction of adverse outcomes in the year after hospitalization.

To test the effectiveness of the intervention on mobility, LSA scores and changes in LSA scores for patients in the control condition will be compared to LSA scores from patients in the intervention condition. Mixed models will be used to examine and compare mean LSA over time with time considered a random effect. An indicator term will be included to denote whether the observation was from the period prior to or after implementation of the intervention. The indicator term will represent whether there was an overall difference in the LSA during the intervention period versus the pre-intervention period, and its statistical significance in the model for each outcome will be the primary test of whether the intervention had a significant effect on LSA. A priori we believe that an exchangeable covariance structure will be most suitable for the data. For the secondary hypotheses, i.e., self-reported outcomes for driving (yes/no) and ability to walk ¼ of a mile (yes/no), generalized linear mixed models (GLMMs) that account for repeated measures will be used to compare pre- and post-intervention periods. Given the randomized nature of the study design, adjustment for potential confounders will likely be unnecessary as they are expected to be equally distributed between the groups. However, as noted above, we will confirm that any potential confounders, e.g., age, length of hospitalization, and disease burden score (Cumulative Illness Rating Scale-Geriatrics[57]) are indeed similar between the pre- and post-intervention periods and, should any not be, we will adjust for them in the statistical models provided they are not deemed to be consequences of the intervention.

To test the effectiveness of the intervention on adverse outcomes, we will use mixed models and GLMMs to model changes in functional decline (ADLs and IADLs), number of ED visits, and number of hospitalizations with an approach similar to the above. To test the effectiveness of the intervention on nursing home admission and death, we will use Cox proportional hazards models

to compare the time from hospital discharge to the time to event between the MP and UC groups, adjusting for potential covariates. Due to the potential interrelationship between risk of nursing home admission and risk of death, competing hazards will be used to model the time to occurrence of the first event. Right censoring of participants will occur with dropout due to any reason (death, loss to follow-up, conclusion of the study period, etc.). Multiple imputation techniques will be used to offset potential bias due to missing predictors in the models.

AIM 2: To identify characteristics which modify the effect of the mobility intervention on recovery to pre-hospital mobility status or better and reduction of adverse outcomes in the year after hospitalization. Potential effect modifiers include pre-admission ADL and LSA, cognitive function, level of in-hospital mobility, depressive symptoms, and social support. For continuous variables, e.g., activities of daily living score, overall distributions will be reviewed and categorized as low, medium and high based on tertile cut-points. To test for significant effect modification, single second level interaction terms, combining pre-/post-intervention indicators with potential effect modifiers, will be added to mixed models. Resulting F-tests for the interaction terms will be reviewed and p-values <0.05 identified as significant. Examination of potential effect modifiers based upon their tertile groupings not only allows for the identification of significant factors but also has the added benefit of examining factor's dose effect, providing additional internal validity to the analysis.

Implementation/Dissemination: This research holds great promise as we begin to define a standard of care for hospital mobility, which is currently lacking. With the success of this proposal we will have demonstrated the effectiveness of this program on mobility and adverse outcomes after hospitalization as well as identified patient characteristics that modify the intervention effects. It is critical to develop a robust evidence base for mobility during hospitalization as hospitals all across the country and the world are currently grappling with the issue of low mobility and its consequences. As noted by a recent review by Kalisch, "Interventions and policies that increase inpatient mobilization need to be developed, tested and put into practice." [6] This study would provide the type of data needed to develop VA policies and a standard of care for safe mobility for all hospitalized veterans. Future steps would include a multicenter trial to assess the generalizability of the program followed by dissemination of the mobility program to VA hospitals across the nation.

We propose a multi-level dissemination plan designed to target research findings to several different audiences. First, we will use traditional dissemination through the publication of high-quality peer-reviewed articles throughout the study's conduct, including empirical, methodological, and theoretical manuscripts. Manuscripts summarizing results will be targeted to (1) management-oriented journals (e.g., Journal of Nursing Management); (2) clinical nursing, physician and physical therapy journals (e.g., Journal of Hospital Medicine); and quality improvement and implementation oriented journals (e.g., Journal of Nursing Care Quality, Implementation Science). We will deliver presentations to research and policy-oriented audiences. It is imperative also that we bring early results to VA-specific audiences including VISN, facility, and nursing leadership; nurse managers, rehabilitation managers and frontline staff. We will use written, oral and web-based dissemination strategies to bring early results to these stakeholders. Examples include reporting through relevant approved Central Office publications (e.g., newsletters, web-based material), as well as reporting to groups of stakeholders who

regularly meet by phone (e.g., Nurse Executives). Finally, we will seek out opportunities to present our findings to frontline staff (e.g., through quality collaboratives or other activities coordinated through the National Center for Patient Safety).

Limitations of the Study: Maintenance of blinding is a potential limitation. It is possible the Assessor will observe the veteran walking, or will be told information about the group assignment by the patient or healthcare provider. We will address this concern in three ways. First, we will regularly encourage the participant not to share their group assignment with the Assessor. In addition, we will provide **all** patients with identical StepWatches®. Lastly, after the patient completes the study, we will ask the Assessor to identify which group they believe the patient was assigned. There is a risk that the walking intervention will lead to adverse outcomes, such as an increase in falls. Our previous pilot study of a mobility intervention demonstrated the intervention group experienced no falls while three falls occurred among patients in the control group. However, despite these positive results, we will make every effort to provide a safe environment for the increased mobility we are proposing by encouraging patients to both recognize and not exceed their mobility capabilities, and by emphasizing that help is readily available and should be requested if needed. We will also provide ambulatory devices to persons who request them and demonstrate safe use. The selection of a Data Safety and Monitoring Board will be critical to help us achieve this objective of the study, i.e. being falls and injury free. (See Human Studies Section for details).