A RANDOMIZED CONTROLLED TRIAL OF REGIONAL VERSUS GENERAL ANESTHESIA FOR PROMOTING INDEPENDENCE AFTER HIP FRACTURE (REGAIN TRIAL)

STATISTICAL ANALYSIS PLAN

Medical, Functional, and Utilization Outcomes up to 60 Days
Version date: 06/18/2021

Principal Investigator	Mark D. Neuman, MD, MSc University of Pennsylvania Department of Anesthesiology and Critical Care 308 Blockley Hall, 423 Guardian Drive Philadelphia, PA 19104 Email: neumanm@pennmedicine.upenn.edu
Study Statistician	Susan Ellenberg, PhD Department of Biostatistics, Epidemiology and Informatics University of Pennsylvania Perelman School of Medicine 611 Blockley Hall, 423 Guardian Drive Philadelphia, PA 19104-6021 Email: sellenbe@pennmedicine.upenn.edu
Study Statistician	Rui Feng, PhD Department of Biostatistics, Epidemiology and Informatics University of Pennsylvania Perelman School of Medicine 209 Blockley Hall, 423 Guardian Drive Philadelphia, PA 19104 Email: ruifeng@pennmedicine.upenn.edu
Funding sponsor	Patient Centered Outcomes Research Institute 828 L St., NW, Suite 900 Washington, DC 20036 Phone: (202) 827-7700 Fax: (202) 355-9558
Grant Number	PCS 1406-18876
ClinicalTrials.gov Number	NCT02507505

I. Introduction

This document describes the statistical analysis plain for mainline analyses for the REGAIN Randomized Trial (Regional vs General Anesthesia for Promoting Independence after Hip Fracture, NCT02507505) related to the primary study outcome and secondary/exploratory outcomes related to medical complications, functional recovery, and health care utilization up to approximately 60 days after randomization. Analyses related to other secondary and exploratory outcomes are described in separate documents. For information on the background, design, and methods of the REGAIN study, please refer to the full study protocol.

II. Quality Control Prior to Unblinding

Prior to unblinding, we will review the distributions of each variable to be included in the analysis, aggregated across treatment arms, to identify any outlying values that may be inaccurate and should be checked. To the extent possible, inaccuracies will be resolved and the database updated with the correct values. Data that are clearly incorrect but cannot be corrected will be excluded from the analyses. Data that are unusual but not impossible, and cannot either be verified or corrected, will remain in the analysis.

III. Baseline and Intraoperative Data

Baseline demographic and clinical variables will be examined to evaluate general trends and determine whether there are any notable imbalances that may lead to secondary adjustments. Continuous variables will be summarized through standard measures of central tendency and spread including means, medians, standard deviations and interquartile ranges (IQRs). Frequency distributions will be calculated for categorical variables.

Intraoperative data, including surgery and anesthesia characteristics, will be summarized similarly.

IV. General Principles

All analyses (primary, secondary, exploratory) will adjust for the 2 binary stratification factors, namely sex, fracture type (intracapsular, extracapsular), and country (U.S., Canada sites). Sex, facture type, and site were the original stratification factors for randomization assignment, but with 46 participating sites, some with very few enrolled subjects, it will not be possible to include site in the stratified analysis. Because there might be systematic differences in health care and insurance use between Canada and U.S. sites, we will add country as an additional stratification factor. We will perform descriptive analyses of results by site to see whether there appear to be any major site-specific effects. Analyses will follow a modified intention-to-treat principle, with sensitivity analyses performed to assess the potential impact of missing values and noncompliance. Subjects who had randomization codes generated due to purely technical errors in operating the database will be excluded from any analysis. Subjects who died before receiving either study treatment will also be excluded from any analysis. Subjects randomized and later found not to meet eligibility criteria will be included in the primary analysis to avoid potential bias due to differential assessment of eligibility. Sensitivity analyses excluding these subjects will be conducted. Subjects who were randomized into the wrong stratum will be analyzed as per-randomization strata. All analyses will use 2-sided tests and an overall significance level of 0.05 as the threshold for statistical significance. We will report nominal p-values for all analyses.

V. Primary Outcome Analysis

The primary endpoint of the study is whether or not the subject is alive and capable of walking 10 feet or across a room without human assistance at 60 days post-randomization. As per the study protocol, the day 60 assessment will be performed within a window of 31-90 days post-randomization, although every effort will be made to make the assessment as close to day 60 as

possible. Results of assessments performed outside the 31-90 day window will be considered as missing data. The treatment effect will be evaluated using the Mantel-Haenszel (MH) test for difference in proportions, stratified by sex, fracture type, and country. The MH relative risk (RR) of spinal anesthesia (SA) vs. general anesthesia (GA) will be reported. *The Breslow-Day test* (Breslow and Day, 1980) will be used to test homogeneity of the RRs across strata. If there is statistically significant heterogeneity, we will report separate RRs within strata.

The walking ability component of the primary outcome will be assessed by a telephone interview with the subject conducted as close to day 60 as possible; information obtained during a 31-90-day window after randomization will be considered acceptable for analysis. To assess walking ability, respondents were asked to indicate whether on average in the 7 days prior to the interview they used/received human assistance to ambulate. Individuals indicating that they did not ambulate at all over the 7 days preceding the interview were coded as having met the study outcome. Vital status will be assessed at the day 60 telephone interview; for patients who cannot be reached for the 60-day telephone interview, vital status at 60 days will be assessed via telephone follow up at subsequent time points. Additional information on deaths prior to day 60 will be obtained from site Adverse Event reports and via a National Death Index search for patients who were randomized at US sites and for whom sufficient identifying information is available. Patients who are known to have died by day 60 will be assessed as having met the primary outcome, regardless of walking status assessments prior to that point. Those known to be alive at day 60 but without a response to the question about capability of walking 10 feet will be considered to be missing the primary outcome. In cases where patients were unable to provide their own responses to the primary outcome assessment (e.g., due to dementia or acute illness), data were collected from alternate (proxy) respondents as described in the main study protocol. Subjects for whom information on both components of the primary endpoint are missing (i.e., patient has become lost to follow-up so that neither survival status nor walking ability is known) will be excluded from the primary analysis, but a variety of sensitivity analyses, described below, will be conducted to assess the potential influence of these exclusions on the comparison of outcomes.

V.1. Sensitivity Analysis for Missing Data

Baseline characteristics of patients with missing primary outcome will be compared with the baseline characteristics of those included in the analysis to assess potential for bias based on these missing values.

The primary analysis will use complete data, excluding individuals without outcome data. To assess the potential for bias incurred by ignoring the missing data, we will conduct a sensitivity analysis under the assumption of missing at random (MAR) using the inverse-probability-weighting (IPW) method (Hogan and Lancaster, 2004). In the calculation of the M-H RR and test statistic, each subject will be weighted by the inverse probability of being a complete case. We will use the non-missing rate of the primary outcome within each treatment and each stratum to estimate this probability.

V.2. Sensitivity Analysis for Noncompliance

To assess the potential impact of non-compliance on the study outcomes, we will use a stratified-adjusted instrumental variable approach to estimate the average effect of spinal anesthesia vs general anesthesia on the primary endpoint for those who actually receive the assigned treatment (Matsui, 2005). Randomization assignment will serve as the instrumental variable because it is designed to be independent of unobserved confounders. The study population is assumed to include individuals who receive the assigned treatment (i.e. “compliers”), individuals who would receive general anesthesia regardless of their study assignment (“always users of general anesthesia”) and individuals who would always receive spinal anesthesia regardless of the study

assignment (“always users of spinal anesthesia”). The observed effect of treatment assignment on compliers is attenuated by the presence of “always users” (i.e. those who “cross over” if they are not assigned their preferred treatment). To correct for this attenuation, an estimated risk of the primary outcome among compliers will be calculated for individuals assigned to spinal anesthesia and general anesthesia separately, within each stratum (Greenland, 2000). The Mantel-Haenszel relative risk for compliers will be calculated as the ratio of the weighted summations of the adjusted risk over all strata. This estimate maximizes asymptotic efficiency even if the compliance rate varies across randomization stratum (Matsui, 2005).

VI. Secondary Outcome Analyses

VI.1. 60 day mortality. This component of the primary endpoint will be assessed via site reports of deaths and at the day 60 telephone interview, and will be analyzed as a binary outcome using the MH method, stratified by sex, fracture type, and country. Patients who are unable to be assessed directly or by proxy at the 60-day telephone interview but who are known to be alive at later time points will be categorized as alive at day 60. Additional information on deaths prior to day 60 will be obtained via a National Death Index search for patients who were randomized at US sites with sufficient identifying information and for whom no follow-up at or beyond day 60 is available. The Breslow-Day test will be used to assess homogeneity of the RRs across strata.

VI.2. Ability to walk 10 feet without human assistance at 60 days post-randomization. This component of the primary endpoint will be analyzed as a binary outcome, using the MH test stratified by sex, fracture type, and country. Participants who died prior to completing the 60 day interview or had no informative answer to the ambulation assessment will be excluded from the analysis.

VI.3. New-onset delirium. This outcome will be determined by baseline and post-operative 3D CAM tests. Subjects will have a baseline 3D-CAM assessment prior to randomization and assessments on days 1, 2 and 3 post-surgery, with fewer assessments if discharged, withdrawn, or died prior to day 3. Only subjects with a baseline 3D-CAM assessment but no delirium will be included in the analysis. Subjects with a positive response on any post-surgery assessment will be classified as positive for new onset delirium. Subjects who had 1 or more post-surgery assessments with none positive for delirium will be classified as negative for this outcome. Subjects without any post-surgery 3D-CAM response will be excluded from this analysis. This binary outcome will be analyzed using the MH method, stratified by sex, fracture type, and country. The Breslow-Day test will be used to assess homogeneity of the RRs across strata.

VI.4. Days to hospital discharge. Days from randomization to hospital discharge or death, censored at post-randomization day 30, will be analyzed using a competing risk Cox regression model, adjusted for sex, fracture type, and country (Fine and Grey 1999). The cumulative incidence of hospital discharge will be compared between study arms. The proportional hazards assumption will be assessed using both survival graphs and formal statistical tests of zero slopes in the Schoenfeld residuals and if the proportional hazards assumption is violated, advanced techniques in survival such as incorporating time-dependent covariate effects will be applied (Grambsch and Therneau, 1994; Hess, 1995). Kaplan-Meier curves displaying times to discharge and accounting for competing death risk by study arm will be presented (Pepe and Mori 1993). The interaction between treatment and country will be tested in the Cox model and if the p-value for the interaction test is 0.2 or less, separate analyses by country will be presented.

VII. Exploratory Analyses

For all exploratory outcomes, summary statistics (frequencies and percentages for categorical variables, means and standard deviations, or medians and IQRs for continuous variables, frequencies and median time to event for time-to-event variables) will be presented by arm. If there are sufficient events for analysis of binary outcomes, the Mantel-Haenszel relative risk, stratified by sex, fracture type, and country, will be estimated.

VII.1. Inpatient major morbidity and mortality. Study site personnel collected data on any of the below events occurring prior to death, discharge, or post-randomization day 30 via medical record review using outcome definitions listed in the REGAIN manual of procedures; additional information on postoperative acute kidney injury was obtained by comparison of site-reported pre-randomization creatinine vs peak post-operative creatinine values where available. Except where specified below, all morbidity and mortality endpoints will be analyzed as binary data.

a. In-hospital death
b. Myocardial infarction
c. Troponin elevation without diagnosed myocardial infarction
    i. Note: Information on peak postoperative troponin values was collected from REGAIN patients based on laboratory data obtained as a component of routine care. Information on isolated troponin elevation (i.e. without a diagnosis of myocardial infarction) will be reported for individuals not otherwise identified as having experienced a myocardial infarction.

       Postoperative troponin data will be presented as a categorical variable with the following categories: (1) postoperative troponin value measured, peak value above the lower limit of detection at the enrolling site; (2) postoperative troponin value measured, peak value below the lower limit of detection; (3) postoperative troponin value not measured. This variable will be analyzed as a categorical variable using a multinomial logistic regression, adjusted for sex, fracture type, and country.
d. Non-fatal cardiac arrest
e. Stroke
f. Pneumonia
g. Pulmonary embolus
h. Pulmonary edema
i. Unplanned postoperative intubation or need for mechanical ventilation
j. Acute kidney injury
k. Surgical site infection
l. Urinary tract infection
m. Any postoperative transfusion
n. Any return to OR
o. Need for critical care admission
p. Fall within 12 hours of anesthesia

VII.2. Time to first in-hospital ambulation after randomization (days).

Time from randomization to first ambulation, censored at hospital discharge, will be compared using a Cox proportional hazards regression model, adjusted for sex, fracture type, country, and use of cane or walker at baseline. Kaplan-Meier curves displaying times to first ambulation by study arm will be presented.

VII.3. Discharge location.

The location is a categorical outcome with five categories: (i) home and independent living retirement home; (ii) nursing home; (iii) rehabilitation or acute care hospital; (iv) other; and (v) death. If there are sufficient subjects in each category, this outcome will be analyzed using a multinomial logistic regression, adjusted for sex, fracture type, and country.

VII.4. Additional 60-day outcomes.

a. Time to death: Days from randomization to death, censored at day 60, will be compared using a Cox proportional hazards regression model, adjusted for sex, fracture type, and country. Kaplan-Meier curves displaying times to death by study arm will be presented. For those with missing data on vital status at day 60, we will search the National Death Index. If death month and year are available but the day is missing, day 15 will be used as an estimate of the date.

b. Worsened walking ability: Subjects are coded as "worsened" if at day 60 they needed a level of assistance in walking greater than prior to surgery, or if they could not walk at all, or if they died prior to day 60. Specifically, those who could walk without human assistance at baseline are considered to have worsened if at day 60 they required assistance to walk (e.g. cane, walker or human assistance), could not walk at all, or died prior to day 60; Those who required a cane or walkers at baseline are considered worsened if they needed human assistance to walk or could not walk at day 60, or if they died prior to day 60.

c. Death or transition to institutional residence at 60 days: We will ask patients their residence locations before hospital admissions. We will also assess a location of residence on or around post-randomization day 60 via blinded telephone interview with the patient or their proxy. The analysis for this endpoint will include all individuals who were admitted from home or retirement home. Individuals who died before 60 days or reported change of residence location (including nursing home, acute rehabilitation, acute care hospital, and other) at the 60 day interview will have "positive" responses; individuals remained at home or independent living retirement home will have "negative" responses. Subjects without a valid admission location, or a 60-day vital status and valid residence location, will be excluded from the analysis. This outcome will be analyzed as a binary variable.

d. Disability score (optional): The disability score will be calculated based on the answers to Questions 1-12 in the World Health Organization Disability Assessment Schedule 2 (WHODAS 2.0) Short Version. When only one item is missing a value, the mean score of the other items will be the assigned score to the missing item. When more than one item is missing, a summary score will not be calculated. Of note, for subjects who were not able to provide their own responses to the WHODAS assessment, data was collected from an alternate (proxy) respondent as described in the main study protocol. Proxy scores will be used if there is no patient score. If scores from both patient and proxy are available, the one with fewer missing/NA items will be used. This variable will be analyzed using a linear regression model, adjusted for sex, fracture type, country, age, and baseline disability score. This score will be analyzed with linear regression models, stratified by sex, fracture type, country, and baseline value if relevant. The normality assumption will be assessed and if it is violated, normalization transformation will be applied before the analysis.

VIII. Subgroup Analyses

We will evaluate treatment difference in the subgroups defined by the following stratification variables:

a. Sex: Male vs. Female, defined as per stratification
b. Fracture type: Femoral neck vs intertrochanteric/subtrochanteric, defined as per stratification
c. Country of enrollment: Canada vs US
d. Need for assistive devices to ambulate prior to fracture: present vs. absent
e. Age category: “oldest old” (85 years and over) vs all others
f. History of Alzheimer’s disease or other dementia: present vs absent
g. History of chronic obstructive pulmonary disease or asthma: present vs absent
h. History of coronary artery disease or prior myocardial infarction: present vs absent
i. Community residence prior to fracture: home or independent living retirement home vs nursing home, rehabilitation or acute care hospital vs other.

The first step will be to test for interaction between treatment and variable defining the subgroups. Analysis within subgroups will be performed for any variable showing an interaction yielding a p-value of 0.20 or less. Subgroup analyses will be analyzed similarly as for primary analysis.

IX. References:

1. Breslow NE, Day NE. 1980. Statistical Methods in Cancer Research: Vol. 1 - The Analysis of Case-Control Studies. Lyon, France, IARC Scientific Publications.
2. Horvitz DG, Thompson DJ. 1952. Generalization of sampling without replacement from a finite universe, Journal of the American Statistical Association, 47: 663–685.
3. Wooldridge JM. 2007. Inverse probability weighted estimation for general missing data problems. Journal of Econometrics, 141(2): 1281–1301.
4. Hogan J, Lancaster T. 2004. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. Statistical Methods in Medical Research, 13, 17 - 48.
5. Matsui S. 2005. Stratified analysis in randomized trials with noncompliance. Biometrics, 61: 816-823.
6. Greenland S. 2000. An introduction to instrumental variables for epidemiologists, *International Journal of Epidemiology*. 2000; 29(4): 722–729.
7. Grambsch PM, Therneau TM. 1994. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika, 81: 515–526.
8. Hess KR. 1995. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. Statistics in Medicine, 14: 1707–1723.
9. Fine JP, Gray RJ. 1999. A proportional hazards model for the subdistribution of a competing risk. Journal of American Statistician Association, 94: 496–509.
10. Pepe MS, Mori MK.1993. Marginal or conditional probability curves in summarizing competing risks failure time data? Statistics in Medicine, 12: 737–751.
11. King G, Zeng L. 2001. Logistic Regression in Rare Events Data. *Political Analysis,* 9: 137-163.