Original CSM-S Statistical Analysis Protocol

Developed in 2013 by Dr. James Dziura and Dr. Zoher Ghogawala

After funding was received from NIH and PCORI we published the planned CSM-S study design, rationale, and protocol.[1]

Preliminary observational data showed mean 1-year SF-36 PCS difference scores of 8.7 for ventral surgery compared with 4.0 for dorsal procedures (SD between 10 and 12) and correlations of baseline with 1-year SF-36 PCS between 0.6 and 0.7. We calculated a total sample size (2:3 ventral-dorsal randomization) required to detect a 5-point difference between ventral and dorsal groups.

A minimum sample size of **137** patients provides at least 90% power (see Tables below):

Surgery (N)	Pre-op	Post-Op	Correlation between Pre- and Post-Op	Difference
Ventral (45)	35.5 ± 10.3	44.2 ± 11.7	0.64	+ 8.7 ± 8.2
Dorsal (70)	35.8 ± 11.3	39.8 ± 11.6	0.66	+ 4.0 ± 9.5
Dorsal Fusion (42)	35.0 ± 11.7	39.6 ± 12.4	0.65	+ 4.6 ± 10.0
Laminoplasty (28)	37.0 ± 10.9	40.1 ± 10.6	0.67	+ 3.1 ± 8.8
All patients (115)	35.7 ± 10.9	41.5 ± 11.8	0.64	+ 5.8 ± 9.7

Difference / SD	Correlation	80% Power / 5% Type I error (2-sided)			90% Power / 5% Type I error (2-sided)		
		$N_{Ventral}$	N_{Dorsal}	N	$N_{Ventral}$	N_{Dorsal}	N
5 / SD = 10	0.60	38	56	**94**	50	74	**124**
5 / SD = 10	0.65	40	59	**99**	53	78	**131**
5 / SD = 10	0.70	42	62	**104**	55	82	**137**

A minimum sample size of 137 across both study arms was inflated by 15% to accommodate anticipated attrition during follow-up, for a final accrual goal of 159 randomized.

Primary analyses will include all subjects randomized using an intent-to-treat approach. The primary endpoint is the 1-year change in Physical Component Summary of the SF-36 (SF-36 PCS) at one year. A likelihood-based analysis using a mixed model will be used to compare SF-36 PCS between groups. This model will adjust for baseline SF-36 PCS, as well as study surgeon using a random effects model. All time points will be included in the model, and each subject will contribute data for the time points at which they were assessed. The model will enable a statistical comparison between treatment groups at each time point, though the comparison at the one-year time point will be the primary analysis. The primary advantage of the mixed model, when compared to commonly used methods such as complete case analysis and single imputation (e.g. last observation carried forward), is its flexibility in handling missing data. This analysis will assume that missing data occurs at random (i.e. the missing data value can be dependent on observed data, but independent of unobserved data). The inclusion of all follow-up time points in the model as well as covariates identified to

be associated with withdrawal will assist in meeting this assumption and minimizing the risk of bias. Although the assumption for missing data is weaker under the likelihood based analysis compared to complete case analysis, a non-ignorable missing data mechanism is possible. Sensitivity analysis using selection and pattern mixture models will be employed to evaluate the robustness of conclusions to the missing at random (MAR) assumption.

Revised Statistical Analysis Plan

June 11, 2019

The following plan was developed by Dr. Norma Terrin, Dr. Karen Freund, Ms. Janis Breeze, and Dr. Zoher Ghogawala prior to the review of any outcomes data from the CSM-S trial.

A revised sample size calculation was performed from preliminary data, partway through the trial, using an estimated within-group standard deviation of 9 points. The revised sample size estimate was consistent with the original number of 159 patients.[1]

Primary analysis will compare change in 1-year SF-36 PCS outcomes for patients as randomized. We will perform an unadjusted analysis because we determined that the baseline characteristics were not different between ventral and dorsal groups. We will not perform any specific analysis for missing data because the 1-year follow-up was 95%.

Since those randomized to a dorsal surgery included both DF and DL, a pre-specified secondary analysis of patients will compare 1-year change in SF-36 PCS among the ‘as treated’ groups (DF versus VF, DL versus DF, DL versus VF). This

secondary analysis is pre-specified - ‘as treated’ (placing cross-overs into their actual treatment cohort) and reflects non-random treatment assignment in the dorsal arm (DL versus DF, as selected by the treating surgeon). Continuous outcomes (change scores for SF-36 PCS, NDI, mJOA, EQ-5D) will be compared using ANOVAs for differences in means, with 95% confidence intervals. Categorical outcomes (risk of complications; health resource utilization) will be compared using chi-square tests, with 95% confidence intervals for difference in proportions. All testing will be two-sided with $\alpha=0.05$.

Revised Statistical Analysis Plan – 2 year outcomes data

July 12, 2020

The following plan was developed by Dr. Norma Terrin, Janis Breeze, and Dr. Zoher Ghogawala prior to the analysis of any 2-year outcomes data.

Primary analysis will include all subjects randomized using an intent-to-treat approach. The primary endpoint is the *1-year change* in Physical Component Summary of the SF-36 (SF-36 PCS) at one year.

Analysis Plan for Year 2 Outcomes

Year 2 outcomes for SF-36 and other continuous outcomes will be analyzed using linear mixed effects models (lme) that include both first and second year outcomes. The outcomes will be change from

baseline. The purpose of the lme is to address missing second year outcomes. Since the year 2 assessments were missing mainly for logistical reasons (at 2 sites, research staff neglected to follow up with participants during specific time intervals), the missing at random (MAR) assumption of the lme is valid. The model will include treatment group, time period, and group*time interaction. To address within-subject correlation, patient id will be a random effect. There will be no adjustment for surgeon, since some surgeons had few patients, and no adjustment for baseline variables, since these were similar between groups. The treatment effect at year 2 will be tested using model contrasts. The year 1 treatment effect will be tested similarly, to confirm previous analyses obtained by t-test and anova. The models will be used to compare the 2 randomized groups and the “as treated groups.” We will not address multiple comparisons for the “as treated” analyses because they are considered to be exploratory.

Analysis Plan for Return to Work

The return to work analysis will include all time periods up to 12 months (1, 3, 6, and 12 months). Groups will be compared via Kaplan-Meier plots and log-rank tests. To address multiple comparisons when comparing the as-treated groups pairwise comparisons will be performed at p=0.025 to account for 2 comparisons (DL vs DF and DL vs VF).

References

1. Ghogawala Z, Benzel EC, Heary RF, et al. Cervical spondylotic myelopathy surgical trial: randomized, controlled trial design and rationale. *Neurosurgery.* Oct 2014;75(4):334-346.