**Project:** Project 2, BIOS 6623

**Report:** VA Surgery Database Review

**Investigator:** Rachel Johnson

**Date:** 1 November 2017

**Introduction**

The goal of this project is to evaluate 30 day mortality rates among 26,518 patients who have undergone heart surgery 44 VA hospitals in the most recent six-month period. Certain demographics were collected among these patients, including hospital code; the six month period in which they had their surgery; procedure type—CABG or valve surgery; height; weight; BMI; ASA—a code for the patient’s condition at the start of surgery; albumin levels; and 30 day mortality. Observed and expected mortality rates were used to determine which hospitals’ mortality rates were clinically and/or statistically unusual and warranted a site visit. The purpose of these site visits would be to either identify potential problems if the rates were unusually high or to perform a site visit to determine if there is anything to learn from in order to improve outcomes at other hospitals if the rates were unusually low.

**Methods**

I was assigned to analyze VA Data 2.

There were originally 26,420 individuals from 44 hospitals in the data set. 2 individuals were removed because they underwent surgeries other than valve surgery or CABG surgery, so they should not have been included in the data set.

There were issues with weight entry in hospitals 1-16 in the most recent period, as weights were entered in kilograms. Other BMIs in the data set were plotted against the calculated BMIs, and any incorrectly entered or calculated BMIs were corrected individually.

The variable ASA, which is an indicator of the patient’s condition at the start of surgery was dichotomized such that ratings of 1, 2, and 3 which indicate better health were in one category, while ratings of 4 or 5 were in another category, which indicated worse health at the beginning of the surgery.

Demographic information at the overall level was summarized in Table 1, with continuous variables summarized with mean and standard deviation and categorical variables summarized with frequency and percentages. Frequencies of missing data were also noted for each variable.

13,239 individuals (49.9%) of the total population were missing values for albumin. There were no obvious differences in other data set values between those with albumin values and for those without albumin values as determined through marginal plots and frequency tables that would have explained the missingness.

30 day mortality for individuals was modeled in a logistic regression that accounted for procedure type, ASA, and BMI. This regression was also modeled with albumin as a predictor to compare estimates and to determine if albumin would be a valuable measure to collect for all patients in the future.

The estimates from this model were used to determine fitted values for each individual in the last six month period for which all covariate data was available, which were then averaged by hospital to determine expected mortality rates for each hospital in period 39. Hospital 30 had no individuals with all covariate data available since they were all missing BMI values in period 39, so no expected mortality rates could be calculated for this hospital.

These expected mortality rates were compared to the observed mortality rates for each hospital in this period—excluding hospital 30. If the ratio of observed to expected mortality rates was greater than 1.2

Final written report:

* Methods: Describe the methods used to clean and analyze the data. Justify and explain your data analysis approach (~ 2 pages).  Should be written in past tense and should not include results.  Do NOT include equations.
* Results: Present results for analyses described in the methods (~1-1.5 pages). Use Tables and Figures as appropriate, including in the text the full interpretation of statistical results for the main findings (i.e. point estimates, confidence bounds, p-values, interpretation of results of test).
* Conclusions: Interpret your results (~.5 to 1 page) in context of scientific question(s). Also discuss any limitations to your analysis that may affect interpretation or that require additional consideration by the investigator.
* Reproducible research information: See below (**Reproducible Research**).

|  |  |
| --- | --- |
| **Hospitals (n)** | 44 |
| **Patients undergoing heart surgery (n)** | 26518 |
| **Procedure (n (%))** |  |
| Valve surgery | 5610 (21.16) |
| CABG surgery | 21457 (80.91) |
| **ASA (n (%))** |  |
| 3 or less | 6296 (23.74) |
| 4 or greater | 19558 (73.75) |
| Missing | 664 (2.5) |
| **BMI (mean (SD))** | 28.64 (3.78) |
| Missing (n (%)) | 702 (2.65) |
| **Albumin (mean (SD))** | 4.02 (0.55) |
| Missing (n (%)) | 13239 (49.92) |
| **30 day mortality (n (%))** | 834 (3.15) |