Bridget Balkaran

BIOS 6623 Project 2 – VA Dataset 2

11/1/17

Introduction

The goal of this analysis is to explore mortality in relation to cardiac procedures through data generated at 44 VA network hospitals. Subjects were patients who had undergone one of two specific cardiac procedures at a VA hospital: CABG surgery or heart valve surgery. Data is broken down into six month periods and we explores the observed death rates at each hospital during the most recent six month period (labeled period 39), and compared these values to an overall hospital adjusted death rate to determine if mortality outcomes at each hospital are as expected, have increased, or have decreased. These comparisons will allow VA hospital administration to determine if changes need to be made or if concerns need to be addressed at certain hospitals. Data received includes information on mortality outcomes (mortality within 30 days of CABG or valve surgery), type of cardiac procedure (CABG or valve surgery), height (in), weight (lbs), BMI (kg/m2), ASA (ASA physical classification system), and Albumin (g/dL). This analysis has been conducted with **VA dataset 2**. Missing data plays a large role in this analysis and will be discussed in the results section of this report.

Methods

The main software used in this analysis was SAS 3.6 Enterprise Edition. R version 3.4.1 was used for creating graphs. Data was explored and a few concerns were observed. Two procedures were labeled as procedure 2. One was from hospital code 25 in the last six month period (39) and the other was from hospital code 2 in period 39. These were representative of a cardiac procedure other than the procedures of interest and these individuals were removed from the analysis. There were three BMI values that were outliers. One value at 75.1 kg/m2 was observed in hospital code 18, period 30 and another value was observed at 72.3 kg/m2 and these two values were determined to be data entry errors as per the investigator. Another value at 2.4379 kg/m2 was observed in hospital code 20, period 39. This was determined to be an entry error as the decimal was off one digit, and the value was corrected to 24.379 kg/m2. Another issue with the data was that some of the BMI values seemed abnormally low. This was because, in period 39, hospitals 1 through 16 had entered weight values in kilograms instead of pounds. The VA hospital administrators have been alerted to this so it can be resolved for future reports. Weight values for these hospitals were converted from kilograms to pounds to correct the values and the corrected dataset will be sent back to the investigators. From the corrected weight values, BMI values were calculated and these calculated values were used in the analysis. With further data exploration it was noticed that some categories of ASA had a frequency of 0 and to prevent further issues within the analysis, ASA was dichotomized into a low category with levels 1, 2, and 3, and a high category with levels 4 and 5. A correlation analysis was also executed to examine any relationships between variables. There were missing data for each category, with missing data for procedure, ASA, weight, height, and BMI amounting to around 2-3% of the data missing for each variable. For albumin, 49.93% of the data were missing. Missing data were analyzed using PROC MEANS, PROC MI , PROC FREQ and SGplot.

Data were analyzed using PROC LOGISTIC. Two models were run. The primary analysis was run modeling mortality outcomes with procedure, BMI, and categorized ASA. Adjusted expected death rates were calculated from the regression, adjusting for all hospitals in the past 3 years, including the most recent time period. These expected death rates were then compared with observed death rates at each hospital to determine if the observed rate was similar, higher than, or lower than the expected death rate. A percent change of 20% or more is cause for concern. A second analysis was run with albumin in the model. This observed and expected death rates were compared similar to the primary analysis. Observed and adjusted expected death rates were compared using a ratio of the observed to expected death rates. If the ratio was larger than 1.2 or a 20% change then this would indicate cause to visit that hospital. The rates from the model including albumin were then compared to the rates from the primary model to determine if the model with albumin could be biased.

Results

**Missing Data Analysis:**

The main investigation done on missing data was for the variable albumin, missing around 50% of the data. Missing albumin data was distributed differently for those within ASA category 5. Those with an ASA code of 5 were more likely to have available albumin data. An ASA category of 5 representative of sicker patients may have been an indicator to pursue further tests and therefore have data on albumin. In terms of missing data patterns, albumin data seems to be missing at random (MAR), since the missing albumin data distribution was different with different levels of ASA. Although albumin data is MAR, which typically does not bias results, it may have an effect on expected death rates, as ASA codes are dependent on the health of an individual and this influences their mortality. Among the other variables in the model, albumin seems to have similar distributions of missing data and seemed to be missing completely at random (MCAR).

BMI values seemed have around 0-3% missing data which seemed to be MCAR, except for hospital 30, where all of the BMI data were missing. Upon further examination of hospital 30, it was observed that all values for BMI, Weight and Height variables were missing. This is reflected in the absence of adjusted expected death rates for hospital 30. No inferences can be made on this hospital. Data will need to be collected from hospital 30 before any expected death rates can be calculated. Missing data for dichotomized ASA and for procedure seemed to be missing 0 – 5 % and seems to be MCAR. Variables not included in the analysis were also examined. For Weight and Height values, data seemed to be MCAR, missing at 0 - 5%, except for hospital 30 which was missing all weight and height data.

**Primary Analysis:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | N | No. Missing | Recent six month period (39) | N | No. Missing | Past 3 Years |
| Mortality, n (%) |  | 0 |  |  | 0 |  |
| Yes |  |  | 145 (3.28) |  |  | 868 (3.27) |
| No |  |  | 4279 (96.72) |  |  | 25650 (96.73) |
| Procedure, n (%) |  | 103 |  |  | 549 |  |
| CABG |  |  | 3507 (79.27) |  |  | 20908 (78.84) |
| Valve |  |  | 814 (18.40) |  |  | 5061 (19.09) |
| ASA category |  | 115 |  |  |  |  |
| low \*\* |  |  | 1063 (24.03) |  |  | 6296 ( 23.74) |
| high \*\* |  |  | 3246 ( 73.37) |  |  | 19558 (73.75) |
| BMI, median (Q1, Q3) | 4212 | 212 | 27.12(20.80, 30.49) | 25816 | 702 | 28.50 (25.67, 31.19) |
| Albumin, median (Q1, Q3) | 2240 | 2184 | 3.99 (3.63, 4.36) | 13279 | 13239 | 4.00 (3.64, 4.37) |
| Height\*, median (Q1, Q3) | 4226 | 212 | 65.51 (63.87, 67.23) | 25816 | 702 | 65.49 (63.78, 67.23) |
| Weight\*, median (Q1, Q3) | 4212 | 198 | 165.14 (124.38, 188.87) | 25911 | 607 | 174.31 (154.82, 192.82) |

**Table 1. Characteristics of VA Cardiac Patients.**

**\*\* = low category includes ASA 1,2, and 3. High category includes ASA 4 and 5.**

**\* = not included in model**

Table 1 shows the distribution of variables included in the model. Overall for the last six month period compared to the last 3 years of data, distributions of variables seem to be similar. Mortality percentages, as well as those for procedure and ASA categories have similar values. BMI, albumin, and height also have similar distributions as shown by quantiles. Weight seems to be a lower in the last six month period compared to the last 3 years of data.

Based on the criteria above in the methods section, 17 out of 44 hospitals had higher than expected death rates. These hospitals can be viewed in Table 2. Hospitals that had an observed death rate of greater than 20% are bolded in Table 2. These hospitals (34, 17, 7, 31, 23, 35, 21, 13, 28, 41, 4, 26, 12, 39, 37, 24, 3) had consistent rates across six month periods also, although the frequencies of mortality were much higher (between 11 and 19). The VA may want to look more closely at these hospitals, and see where hospitals can improve on mortality outcomes.

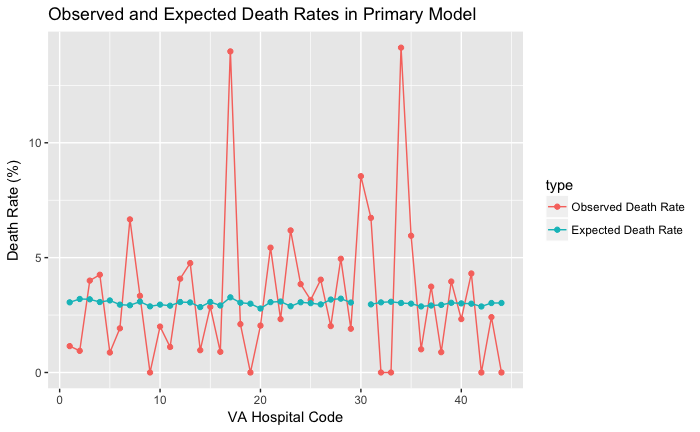
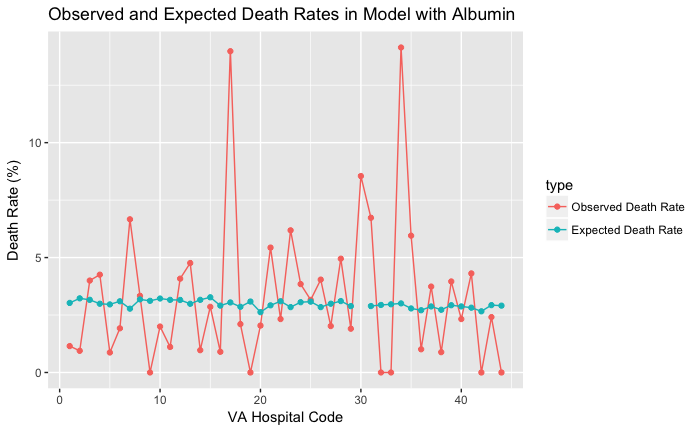
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Hospital | observed death rate | adjusted expected  death rate (Primary Model) | observed:expected  (Primary Model) | adjusted expected  death rate (Model with Albumin) | observed: expected (Model with Albumin) | Percent Change |
| **34** | **0.1414** | **0.0302** | **4.6728** | **0.0301** | **4.7036** | -0.6557 |
| **17** | **0.1397** | **0.0327** | **4.2722** | **0.0305** | **4.5869** | -6.8607 |
| **7** | **0.0667** | **0.0292** | **2.2784** | **0.0278** | **2.4015** | -5.1247 |
| **31** | **0.0673** | **0.0296** | **2.2704** | **0.0289** | **2.3285** | -2.4939 |
| **23** | **0.0619** | **0.0288** | **2.1446** | **0.0284** | **2.1752** | -1.4052 |
| **35** | **0.0595** | **0.0299** | **1.9859** | **0.0279** | **2.1360** | -7.0260 |
| **21** | **0.0543** | **0.0306** | **1.7746** | **0.0292** | **1.8600** | -4.5891 |
| **13** | **0.0476** | **0.0305** | **1.5619** | **0.0298** | **1.5932** | -1.9690 |
| **28** | **0.0495** | **0.0321** | **1.5405** | **0.0311** | **1.5934** | -3.3228 |
| **41** | **0.0431** | **0.0299** | **1.4388** | **0.0282** | **1.5288** | -5.8930 |
| **4** | **0.0426** | **0.0307** | **1.3869** | **0.0299** | **1.4225** | -2.5022 |
| **26** | **0.0404** | **0.0296** | **1.3633** | **0.0284** | **1.4207** | -4.0394 |
| **12** | **0.0408** | **0.0307** | **1.3306** | **0.0316** | **1.2936** | 2.8596 |
| **39** | **0.0396** | **0.0303** | **1.3051** | **0.0293** | **1.3533** | -3.5603 |
| **37** | **0.0374** | **0.0292** | **1.2820** | **0.0287** | **1.3009** | -1.4558 |
| **24** | **0.0385** | **0.0306** | **1.2581** | **0.0306** | **1.2579** | 0.0153 |
| **3** | **0.04** | **0.0319** | **1.2553** | **0.0316** | **1.2654** | -0.8017 |
| 8 | 0.0333 | 0.0308 | 1.0806 | 0.0318 | 1.0483 | 3.0844 |
| 25 | 0.0316 | 0.0302 | 1.0462 | 0.0308 | 1.0245 | 2.1248 |
| 15 | 0.0286 | 0.0307 | 0.9309 | 0.0327 | 0.8732 | 6.6098 |
| 43 | 0.0241 | 0.0302 | 0.7967 | 0.0293 | 0.8223 | -3.1162 |
| 40 | 0.0233 | 0.0301 | 0.7729 | 0.0287 | 0.8091 | -4.4647 |
| 22 | 0.0233 | 0.0309 | 0.7527 | 0.0309 | 0.7502 | 0.3354 |
| 20 | 0.0204 | 0.0276 | 0.7326 | 0.0262 | 0.7777 | -5.8026 |
| 18 | 0.0211 | 0.0304 | 0.6928 | 0.0286 | 0.7367 | -5.9514 |
| 10 | 0.02 | 0.0295 | 0.6769 | 0.0322 | 0.6215 | 8.9314 |
| 6 | 0.0192 | 0.0295 | 0.6510 | 0.0310 | 0.6205 | 4.9361 |
| 27 | 0.0202 | 0.0318 | 0.6361 | 0.0299 | 0.6752 | -5.7858 |
| 29 | 0.0190 | 0.0305 | 0.6253 | 0.0289 | 0.6594 | -5.1665 |
| 11 | 0.0111 | 0.0291 | 0.3821 | 0.0316 | 0.3517 | 8.6577 |
| 1 | 0.0115 | 0.0305 | 0.3763 | 0.0303 | 0.3798 | -0.9122 |
| 36 | 0.0101 | 0.0288 | 0.3512 | 0.0271 | 0.3729 | -5.7957 |
| 14 | 0.0097 | 0.0284 | 0.3410 | 0.0316 | 0.3071 | 11.0303 |
| 16 | 0.0090 | 0.0292 | 0.3088 | 0.0291 | 0.3095 | -0.2311 |
| 38 | 0.0088 | 0.0294 | 0.3008 | 0.0273 | 0.3243 | -7.2443 |
| 2 | 0.0094 | 0.0320 | 0.2948 | 0.0323 | 0.2924 | 0.8253 |
| 5 | 0.0087 | 0.0314 | 0.2772 | 0.0296 | 0.2935 | -5.5777 |
| 9 | 0 | 0.0288 | 0 | 0.0312 | 0 | 8.2689 |
| 19 | 0 | 0.0299 | 0 | 0.0309 | 0 | 3.0800 |
| 32 | 0 | 0.0305 | 0 | 0.0294 | 0 | -3.7805 |
| 33 | 0 | 0.0308 | 0 | 0.0297 | 0 | -3.6038 |
| 42 | 0 | 0.0287 | 0 | 0.0266 | 0 | -7.3712 |
| 44 | 0 | 0.0303 | 0 | 0.0291 | 0 | -3.9092 |
| 30 | 0.0855 |  |  |  |  |  |

Six hospitals had a lower observed death rate than expected. For hospitals with an observed death rate of 0 (Hospitals 9, 19, 32, 33, 42, 44), these hospitals had a higher percentage of CABG surgery, which is considered a less risky surgery than valve surgery but this was also seen in hospitals that had higher death rates. In previous six month periods, these hospitals also had a very low number of deaths amounting to 0-2 deaths per six month period. This trend is consistent with what was observed in the last six month period, and this indicates that although these hospitals are doing more procedures that are less risky, VA administrations may want to look into the policies of these hospitals further, as they may have practices that the VA may want to implement across their network.

Table 2. Results of Observed Death Rates, Adjusted Expected Death Rates, Observed to Expected Death Rate Ratios and Percent Change Between Expected Death Rates in the model with Albumin and the Primary Model.

**Analysis of Model with Albumin:**

The model with albumin seems to show similar results to the primary analysis (Figure 1). The primary model and the model with albumin seem to show the similar values for each hospital for adjusted expected death rates (Table 2). Since the rates we are working with have small variations, Figure 1 allows us to see the spread of observed values compared to expected values. Although the expected death rate for each hospital seems to hover around 3%, observed rates vary greatly from 0 to 14%. From figure 1 we can see hospital 17 and hospital 34 have observed death rates much higher than their expected death rates. To compare the observed to expected values, ratios are used to better view hospitals that may need to be addressed. Figure 2 shows the ratio of observed death rates to adjusted death rates in the primary model and in the model with albumin. The black horizontal line in figure 2 denotes a limit of 20% change in the ratio. Any points above this line have a higher than 20% change between their observed rate and their expected rate and any points below this line have a lower than 20% change between their observed rate and their expected rate. The two different models are also shown in figure 2, and it seems that for those with a higher than a 20% change in the observed and expected death rates, the models do not seem to line up and give the same ratio which could indicate some level of bias. For hospital 17 we can clearly see the model with albumin has a higher change than the primary model. This is also seen in in some of the hospitals closer to the delimiting line, while for hospitals under the line, the points seem to line up almost exactly. We can conclude that for the model with albumin, the higher the percentage change, the more likely we are to observe a divergence in the primary model and the model with albumin. The model with albumin could show biased results here, as higher ratios would indicate greater differences in observed and expected death rates.

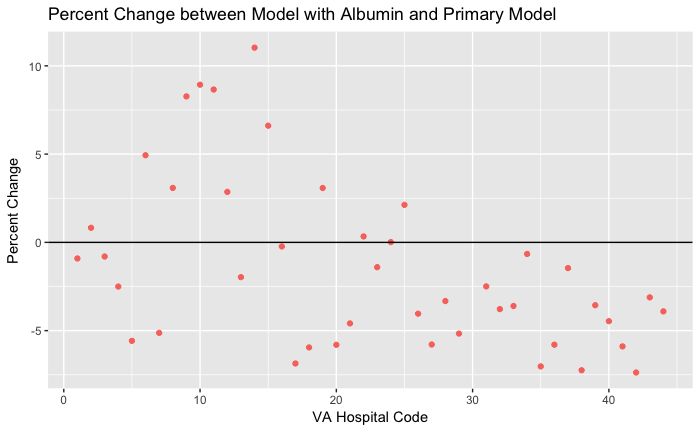


**Figure 1. Observed and Adjusted Expected Death Rates in period 39 in the Primary Model and the Model with Albumin. \*Note connecting lines are not pictured to show trend, but to ease the alignment of points for the reader.**



**Figure 2. Observed to Expected Ratios. Percent Change Comparisons Between Observed and Expected Death rates in the Primary Model and the Model with albumin.**

Lastly, we looked at percent change between adjusted expected values in the Primary Model and those in the model with albumin. Percent change was calculated as (expected death rate with albumin model – expected death rate primary model) / expected death rate in the primary model x 100. Figure 3 shows the distribution of this. The black line indicates no percent change seen. Values above the line indicate positive percent change or percent increase showing that the expected death rate for the model with albumin was larger than the expected death rate for the primary model. Values below the line show that the expected death rate of the model with albumin are smaller than the expected death rates in the primary model. Overall, there seem to be larger expected values from the Primary Model than the model with albumin (more values below the black line). This could mean that for hospitals whose percent change is positive (above the black line) there could be some bias in the model with albumin since these expected death rates yield larger values. From figure 3, we can see that if hospitals that have higher expected death rate values from the model with albumin (values are above the black line), then perhaps hospitals that adjusted for albumin had patients that were more likely to die, resulting in a higher death rates with albumin in the model. This could be due to the fact that sicker individuals (higher ASA values) were more likely to have albumin data which would bias the results away from the null. However this cannot be stated explicitly, since the statistic is not normally distributed and we cannot make statements with confidence without performing a bootstrap confidence interval.



**Figure 3. Percent Change of Expected Death Rate in Primary Model and Expected Death Rate in the Model with Albumin.**

Conclusion:

In conclusion, we can see that in some hospitals, the model with albumin may have biased results since those who are sicker are more likely to have albumin data and those that are sicker are at a higher risk of mortality. In further reports it would be interesting to collect albumin data on all patients to see if expected values are truly different for primary models and models with albumin. In terms of which hospitals should be visited, hospitals bolded in Table 1 may want to be looked at further and hospitals 17 and 34 seemed to have the highest death rates in both models and this may need to be addressed by hospital administration. For hospitals that are doing well in keeping mortality down, VA administration may be interested in visiting hospitals 9, 19, 32, 33, 42, and 44 to see how they are working to improve mortality outcomes after cardiac procedures. These methods may be useful to implement in other VA network hospitals to help keep mortality low. A strength of this study is the availability of past years’ data to calculate adjusted values to have better means of comparing observed and expected death rates. A limitation of this study is that although the observed and expected death rates can be calculated and observed we cannot be sure any change are statistically significant as we cannot assume our population is normally distributed. In further studies, performing a bootstrap to generate confidence intervals would be beneficial to see if changes between observed death rates and expected death rates are significant.

Reproducible Research

Github Path: bios6623-bbalkaraan/Project2/Code

Data file saved in SAS studio: /home/bridgetbalkaran0/my\_courses/BIOS\_6623 Advanced Data Analysis/Project\_2/vadata2.sas7bdat