Investigating the Correlation between Epworth Sleep Scale and Sleep Study Indices, and the Association between Hypoxemia and Cardiac Arrhythmia: UCLA Sleep Lab

Contributors: Karishma Raghuram, Yujin Lee, Chris Thornton,

Pawan (Sine) Polcharoen, Brandon Hao, Kellen Whetstone

STATS141XP Spring 2023

Professor Mahtash Esfandiari

June 11, 2023

TABLE OF CONTENTS

ABSTRACT

This report reviews data obtained from subjects at the UCLA sleep lab over the period of January 1st to March 15th, 2023. Evaluated in this paper are the questions of: 1) whether Epworth Sleep Scale correlates with commonly observed sleep study indices and 2) whether hypoxemia is associated with cardiac arrhythmia. The first question was addressed using non-parametric correlation testing and logistic regression, while in the second question, logistic regression was used to form binary predictions regarding the presence of cardiac arrhythmia. Analysis of the data determined that ESS did not have a statistically significant relationship with any of the typical sleep study indices and that hypoxemia was not associated with cardiac arrhythmia as defined by the binary coding scheme that was used.

INTRODUCTION

Sleep plays a vital role in maintaining overall health and well-being. Quality rest is essential for cognitive function, mental health, and physical rejuvenation. However, in today's fast-paced world, achieving consistent and adequate sleep is often put on the back burner. According to the Centers for Disease Control and Prevention, about 1 in 3 adults reported not getting enough sleep everyday. Furthermore, 50 to 70 million adults in the US are affected by a sleep disorder, with 25 million impacted by obstructive sleep apnea. Children and teenagers also experience sleep deprivation. According to a study by the National Sleep Foundation, more than 65% of people aged 5 to 17 have not reached 8 hours of daily sleep.

Our study analyzes data from patients of the UCLA Sleep Laboratory. The main variable of interest is the Epworth Sleepiness Scale (ESS), a crucial measure of daytime sleepiness. This statistical report aims to explore the factors influencing this ESS measure, and provide meaningful insights into the relationships between these factors, in order to help us better understand how we can improve the average person's sleep. Specifically, we will be looking at correlations between ESS and other factors such as sleep disordered breathing, periodic limb movements, sleep stage, etc. To further our research, we will also look into these relationships stratified by age and gender. Beyond non-parametric correlation, we utilize logistic regression models to determine the key predictors of ESS and the Apnea Hypopnea Index (AHI).

As for our second area of interest, we will be investigating the relationship between oxygen desaturation counts and cardiac arrhythmia. Given data regarding a patient's total oxygen desaturation events below thresholds of 90%, 80%, and 70%, we will be analyzing whether these events correlate with the presence of cardiac arrhythmia.

BACKGROUND

When looking at the Epworth Sleepiness Scale (ESS), a value of 10 or greater is considered problematic. On top of that, a value 16 or greater will indicate severe sleep disorders such as narcolepsy. The question surveyed to obtain the ESS is based on the Likert Scale, which is rated from 0-3 and thus the highest possible score is 24. The scale was initially introduced in 1990 by an Australian psychologist, Dr. Murray Johns. As of 2023, ESS has been available in almost every language and each question is assigned a maximum of three minutes. During the assessment, the respondent is required to rate their tendencies to fall asleep while engaging in eight different activities. As a convenient measuring tool, many doctors and clinical researchers have utilized ESS as well as develop other formulas to identify sleep disorders as well as recognize possible treatments; that being said, the accuracy of said test is somewhat contentious given the design of the survey and lack of efficacy. This ultimately proves problematic as a result of misdiagnosing patients and an abuse of insurance payouts

In diagnosing sleep disorders, the Apnea Hypopnea Index (AHI) is a critical measure, and can be more reasonably deterministic of abnormalities in sleep patterns. AHI is the combined average number of apnea and hypopnea events that occur on an hourly basis. While it has many shortcomings, such as taking the average of both apnea and hypopnea events, which connote different severities in sleep abnormalities (the former being a complete obstruction of airflow, while the latter being a less severe blockage), the AHI in combination with other valuable predictors related to oxygen desaturation, sleep time, etc, prove to be more effective in deriving critical information in sleep disease prognosis.

THE DATA

Our data is collected by the UCLA Sleep Laboratory, under the guidance of Dr. Ravi Aysola. This data includes 400 studies of patient sleep observations, collected from January 1, 2023 through March 15, 2023. This dataset will be utilized in our work modeling for research question 1 and 2. However, for correlation testing for question 1, we will be using a separate data set.

Our main variables of interest other than ESS for the correlation testing are: Apnea Hypopnea Index (AHI), Periodic limb movement (PLM), DESATS LT, Body Mass Index (BMI), and Sleep Efficiency. The Apnea Hypopnea Index (AHI) is defined as the total number of apnea and hypopnea occurrences divided by total sleep time. Values of 5-14/hour are considered mild, 15-30/hour are considered moderate, and $\leq$ 30/hour are considered severe. The movement parameter periodic limb movement (PLM) index is the number of leg movements per hour of sleep. Values equal to or greater than 15 are considered abnormal. In turn, the LEG1+LEG2 Index variables are the number of individual leg movements per hour of sleep. To measure the degree of hypoxemia, we use the variable DESAT, which is the measure of inefficient blood

oxygen during sleep. Specifically, DESATS LT 70, 80, and 90, measure the number of times that SaO2 dips below 70%, 80%, and 90% respectively.

VARIABLES OF INTEREST & DEFINITIONS

- Respiratory parameters:
    - **Apnea Hypopnea Index (AHI)** = total number apnea + hypopnea/total sleep time (TST)
    - **Respiratory disturbance index (RDI)** = total number of apnea + hypopnea + respiratory effort related arousal (RERA)/TST
- Movement parameters:
    - **Periodic limb movement (PLM) index** = total number of PLMs/hr of TST
- **Body Mass Index (BMI):** person's weight in kg divided by the square of height in meters.
- **Epworth Sleepiness Scale (ESS):** measure of daytime sleepiness
    - Patient questionnaire: Epworth sleepiness scale (ESS) greater than or equal to 10 is abnormal
- **REM: Rapid Eye Movement** – sleep stage closest to waking state, involving vivid dreaming
- **Apnea:** Complete pause in breathing. Cessation of breathing for at least 10 seconds.
- **Latency:** The time for someone to fall asleep after the dimming of the lights
- **DESAT:** Measure of inefficient blood oxygen during sleep (oxygen desaturation)
- **DESATS LT 80:** is the number of times the SaO2 dipped below 80%
- **PLM:** Total is the number of leg movements during the recording (vs. the LEG1+LEG2 Index would be the number of leg movements per hour of sleep)
- **Sleep Efficiency:** a percent of time asleep (this could be the % of sleep after sleep latency or % sleep during total recording time—not sure which is used here**)**
- **Hypopnea:** the reduction in ventilation of at least 50% that results in a decrease in arterial saturation of 4% or more due to partial airway obstruction. Essentially, periods of shallow breathing with reduced airflow

CUT OFFS

- **AHI and RDI:**
    - 5-14/hr mild
    - 15-30/hr moderate
    - >= 30/ hr severe
- **Oxygen saturation** <= 88% is considered abnormal.
- **PLMI** >= 15 considered abnormal
- **PHQ-9** score >= 5 is abnormal

SPECIFIC METRICS OF INTEREST

- Total Duration of apnea events/ Total sleep time and total sleep time with hypoxemia (abnormal oxygen level)
- Sleep parameters: arousal index, sleep stage (and 1,2, 3/4, REM)

EXTRA INFORMATION

- **Any INDEX:** is the number of times (events) per hour of sleep that it occurred.
- **Any TIME:** is the number in minutes
- **Any Total:** is the number of events during the study
- **Relationship between Pressure and Sleep:** The pressure increases when awake, while it decreases when sleeping.
- **Note: Cardiac event data is not quantitatively reported.** The technicians enter comments if abnormalities are noted. Will you be able to extract from the field whether a comment was entered E.g. yes/no). This would at least identify the potential abnormal cases. The software records and reports minimum and maximum heart rate, if it is capable of reporting specific time spent with heart rate within certain values.

QUESTION #1

Does an abnormal Epworth Scale score correlate with any abnormalities in the typical indices of sleep studies, e.g. sleep disordered breathing, periodic limb movements, sleep stage, etc.? Stratify this by age and gender.

EXPLORATORY DATA ANALYSIS

The data regarding the Epworth Sleepiness Scale is obtained by sleep technicians during the lab meeting. As a result, the available data is listed under the variable name SleepData: ESS. The ESS data was collected by patients with a value between 0 and 24. Initially, the extraneous values for this variable were deleted in the data because of inconsistency or imprecision of the values. Such data would be highly misleading since values of 888 or 999 will significantly impact the mean ESS value, while also being outside the range of permissible values. Disregarding those invalid values, any value 10 or higher in the scale denotes abnormal sleep patterns. The requisites required to perform our analysis is to incorporate a model that captures the relationship between ESS and predictors including, but not limited to, age, weight, and sleep latency index. In addition to that, data cleaning was involved by removing non-response (0) values in the sleep efficiency index, sleep time, and Time:WAKE, as well as unreasonably large weight or BMI values.

METHODOLOGY/RESULTS

Correlation Testing

Our first step to answer this question was to run correlation tests, looking at the relationship between Epworth Scale score and indices like periodic limb movements, apnea counts, oxygen desaturation, and more. First, we created a correlation plot with all of our variables included, utilizing a heatmap to better visualize the correlations, shown in the figure below. This figure only includes data from those who have abnormal ESS (score above 10).

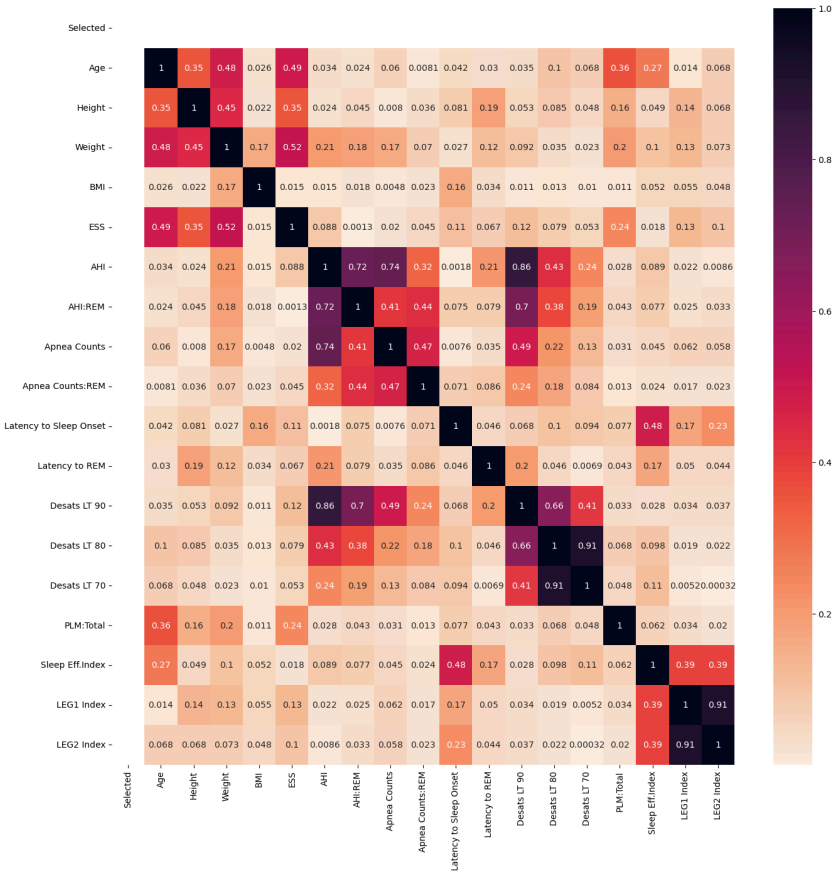

Figure 1 - Correlation Heat Map for ESS > 10

Looking at the heatmap, darker colors represent correlation scores closer to 1, which indicates a stronger positive relationship between the two variables. For example, the variable “LEG1 Index” and “LEG2 Index” have a correlation of .91, meaning a patient that shows movement in one leg is very likely to show similar movement in the other leg as well. The diagonal of the plot shows all 1s, as each variable is perfectly correlated with itself.

As for our response variable of interest, Epworth Scale score (denoted by ESS), we notice it is most closely correlated with age, height, and weight, indicated by the darker colors and

respective correlation scores of 0.49, 0.35, and 0.52. While the correlation scores between ESS and our predictor variables of interest, which include periodic limb movements (PLM: Total), are not observably large based on the heatmap, this does not mean the scores are not significant.

To determine significance of the correlation between ESS and PLM: Total, we can run a Pearson's correlation test, which quantifies the linear relationship between the two variables. We decided to stratify by gender, to better investigate the relationship between the variables, while accounting for differences within subgroups. In this case, gender can be seen as a confounding variable, which means that the relationship between ESS and PLM: Total might be significantly different across genders.

After stratifying our variables and performing the test, we obtained the following results.

PEARSON TEST	Correlation Coefficient	P-value
Female	0.19846	0.06
Male	0.26964	0.002

Figure 2 - Pearson's correlation test results

Observing the above table, we see that the correlation between ESS and PLM: Total is only significant at a level of alpha = 0.05 for males, with a p-value of 0.002. This tells us that there is a significant positive relationship between male's Epworth Scale score and their total leg movements while sleeping.

Furthermore, we looked into the correlation between ESS and Arousals with PLM. Before calculating correlation coefficients, we executed a Shapiro-Wilk normality test on both variables to assess their distribution. The test calculates a test statistic (Calc W), which is a measure of how well the ordered and standardized sample quantiles fit the standard normal quantiles. The statistic will take a value between 0 and 1 with 1 being a perfect match. The low p-values seen in the following tables (Figure 3, 4) indicate that our data isn't normally distributed for both variables, regardless of gender.

Shapiro-Wilk Test: Arousals with PLM	W (test statistic)	P-value
Female	0.50696	2.2×10^{-16}
Male	0.40498	2.2×10^{-16}

Figure 3 - Shapiro-Wilk Test: Arousals with PLM

Shapiro-Wilk Test: ESS	W (test statistic)	P-value
Female	0.39591	2.2×10^{-16}
Male	0.50397	2.2×10^{-16}

Figure 4 - Shapiro-Wilk Test: Epworth Sleepiness Score

Although our variables of interest–Arousals with PLM and ESS–are not normally distributed according to the Shapiro-Wilk test, we ran Kendall, Spearman, and Pearson correlation methods to obtain three correlation coefficient values as depicted in the table below. The non-parametric Kendall and Spearman correlation methods evaluate the strength and direction of a monotonic relationship (either increasing or decreasing) between variables with ranked or ordered data. The assumptions for Spearman are that data must be at least ordinal and the scores on one variable must be monotonically related to the other variable. Spearman correlation is also used when the variables have outliers, as it is less affected by extreme values compared to Pearson correlation. The final method of Pearson measures the strength and direction of a linear relationship between two continuous variables. It assumes that the relationship between the variables is linear and that the data follows a normal distribution. All three methods result in a correlation between -1 and +1, indicating either a negative or positive relationship, with -1 or +1 being a perfect correlation.

Arousals with PLM & Epworth Sleepiness Score	Correlation Method		
	Kendall	**Spearman**	**Pearson**
Female	-0.03325	-0.04330	-0.06613
Male	-0.00323	-0.00355	-0.01450

Figure 5 - Arousals with PLM and Epworth Sleepiness Score Correlations

As we see in the table above, our correlation values for all methods are very close to 0, which indicates a weak, if not near arbitrary, relationship between Arousals with PLM and ESS regardless of gender. The correlation values are all negative, which suggests a negative relationship between Arousals with PLM and ESS.

Modeling For ESS

In addition to correlation, we used a logistic model to help identify which variables in our data set are effective in predicting ESS. Our model predicts binary variable ESS: $0 \leq \text{ESS} \leq 10$ and $10 < \text{ESS} \leq 24$. We chose this binary cut-off as those with ESS greater than 10 are considered abnormal. After removing unnecessary columns and cleaning our data, we were left with the following predictors to be evaluated for our model: Age, Gender, Height, Weight, BMI, AHI,

AHI:REM, Apnea Counts, Apnea Counts:REM, Latency to Sleep Onset, Latency to REM, Desats LT 90, Desats LT 80, Desats LT 70, PLM:Total, Sleep Eff.Index, LEG1 Index, and LEG2 Index.

We then proceeded to split the data into testing (20%) and training (80%) data, for both the predictors and the response variable ESS. We used LASSO regression technique (Least Absolute Shrinkage and Selection operator) for feature selection to help us identify and prioritize the most important predictors among our data set. This technique handles multicollinearity and provides a ranking of feature importance. Using LASSO selection, we removed the following predictors: AHI, Apnea Counts:REM, Desats LT 90, Desats LT 80, PLM:Total, and LEG1 Index. This result is consistent with correlation tests above, as periodic limb movements and AHI are not highly correlated with ESS. With these features removed, we are then able to simplify our model without sacrificing predictive performance.

After subsetting our training data, we fitted a logistic model using our leftover predictors: Age, Gender, Height, Weight, BMI, AHI, AHI:REM, Apnea Counts, Latency to Sleep Onset, Latency to REM, Desats LT 70, Sleep Eff.Index, and LEG2 Index. A classification report of our model revealed an accuracy score of 0.731 overall. The following table displays the summary output of the logistic regression model, which details each predictor along with their model coefficients, standard error, Z-score, and p-value (based on Z-score). Significant predictors with p-values lower than a critical value of 0.05 are highlighted in green.

Predictor	Coefficient	Standard Error	Z score	P-value
Age	-0.3611	0.156	-2.314	0.021
Gender (Male, Female)	0.1576	0.136	1.155	0.248
Height	0.1288	0.147	0.876	0.381
Weight	0.5982	0.177	3.388	0.001
BMI	0.2179	0.171	1.278	0.201
AHI:REM	-0.0552	0.144	-0.384	0.701
Apnea Counts	-0.0721	0.144	-0.501	0.616
Latency to Sleep Onset	-0.2422	0.142	-1.702	0.089
Latency to REM	-0.1419	0.132	-1.072	0.284
Desats LT 70	-0.1240	0.142	-0.871	0.384

Sleep Efficiency Index	-0.3595	0.160	-2.248	0.025
LEG2 Index	-0.0444	0.136	-0.326	0.745

Figure 6 - ESS Prediction Logistic Regression Model Summary

Among the predictors examined, we focus on the most and least significant variables in relation to ESS. The significant predictors are Age, Weight, and Sleep Efficiency Index. Looking at their coefficients, we can understand their effect on ESS;for example, the Age coefficient of -0.3611 is related to the odds of an abnormal ESS score. This finding suggests that older individuals may experience lower levels of sleepiness compared to their younger counterparts. Weight has a coefficient of 0.5982, indicating that as weight increases, sleepiness levels tend to increase as well. It suggests that individuals with higher body weight might be more prone to experiencing daytime sleepiness. Lastly, Sleep Efficiency Index was found to have a significant negative effect on ESS, with a coefficient of -0.3595. This suggests that as sleep efficiency increases, sleepiness tends to decrease. Some non-significant predictors include gender and BMI. Along with the other remaining 7 predictors, these features have no evidence to suggest that they have a substantial influence on ESS.

We proceeded to gather predictions based on our remaining test data, applying the logistic regression model we created above. Along with these predictions, we created an ROC (Receiver Operating Characteristic) curve, which is a graphical representation of a binary classification model's performance. It shows the relationship between the true positive rate (sensitivity) and the false positive rate (1 - specificity) at different classification thresholds. The area under the ROC curve (AUC-ROC) summarizes the overall performance, with a higher value indicating better discrimination.

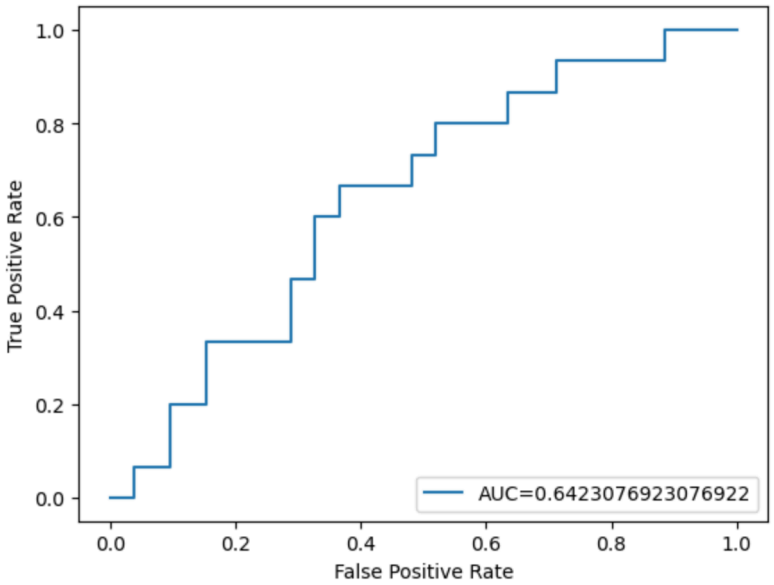

Figure 7 - ESS Prediction Logistic Regression Model ROC Curve

Looking at the ROC curve, we can see that it is curved upwards, which is indicative of a better than average classifier; this is supported by the AUC (Area Under the Curve) value of 0.642. This metric quantifies the overall performance of the classifier. It ranges from 0 to 1, with a higher value indicating better discrimination. An AUC-ROC of 0.5 suggests a random classifier, while an AUC-ROC of 1 represents a perfect classifier. Thus, our model is closer to being a random classifier rather than a perfect one according to the AUC.

Lastly, we created a confusion matrix from our predictions to further evaluate the accuracy, power, and significance of our model. There were 44 true positives, 10 false positives, 8 false negatives, and 5 true negatives. From these values, we obtain an accuracy of 0.731, an error rate of 0.269. The resulting sensitivity is 0.815 and specificity is 0.385. Sensitivity measures how well the model detects true positive cases, while specificity indicates its accuracy in identifying true negative cases.

TEST DATA	Predicted Negative (0)	Predicted Positive (1)
Actual Negative (0)	44	8
Actual Positive (1)	10	5

Figure 8 - ESS Prediction Logistic Regression Model Confusion Matrix (Test)

Modeling For AHI

In addition to predicting ESS, we also fit a logistic regression model on our data to predict AHI. We begin by cleaning the data; this entailed standardizing height and weight measurements, using baseline data, and removing rows with NA values. After the initial exploratory data analysis (EDA), we proceeded to implement LASSO variable selection techniques to determine which predictors were most significant to a predictive model, while minimizing predictive error through optimizing for λ, the tuning parameter. LASSO regression works by scaling all predictor variables so that those which are unimportant to the model go to 0, while those which are essential maintain a quantitative coefficient. Beyond this, we split our data into training and testing data and proceeded to convert our AHI variable into a binary based on abnormal cut-off values. This means we have two levels of prediction: AHI ≤ 30 and AHI > 30.

Using a subsetted version of our data set with the following columns– ESS, Age, Gender, Height, Weight, BMI, AHI, AHI:REM, Apnea Counts, Apnea Counts:REM, Latency to Sleep Onset, Latency to REM, Desats LT 90, Desats LT 80, Desats LT 70, PLM:Total, Sleep Eff.Index, LEG1 Index, and LEG2 Index– we split our data into testing (20%) and training (80%) sets. In choosing predictors of importance derived from our LASSO variable selection, we ascertain that

AHI:REM, Apnea Counts, Latency to REM, Desats LT 90, and Desats LT 70 are key to keep in the model.

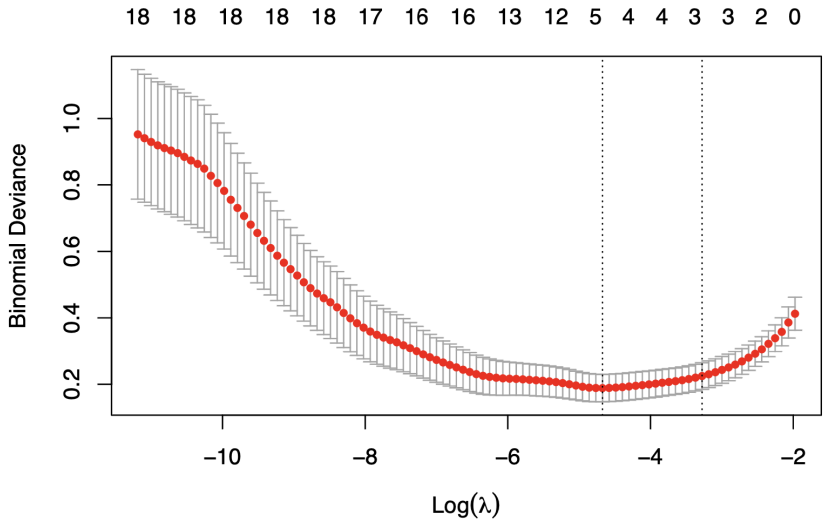

Figure 9 - AHI Prediction Lasso Feature Selection

The graph above displays the cross-validation error according to the log of the regularization parameter (λ). The left dashed black vertical line indicates the optimal value of λ which is the one that minimizes the prediction error (i.e., binomial deviance). The larger λ value, the more coefficients move towards 0 when performing LASSO regression. An optimal λ value prevents the model from overfitting; simply put, this means that the model is predictive not only of the given data that it is formed from, but also new data that may be inputted at a later time.

Predictor	Coefficient	Standard Error	Z score	P-value
(Intercept)	-11.1945	2.626	-4.264	2.01×10^{-5}
AHI REM	0.0685	0.0254	2.696	0.007
Apnea Counts	0.0189	0.011	1.661	0.097
Latency to REM	0.0162	0.006	2.638	0.008
Desats LT 90	0.0420	0.012	3.416	0.001
Desats LT 70	-0.2576	0.101	-2.540	0.011

Figure 10 - AHI Prediction Logistic Regression Model Summary

TRAIN DATA	Predicted Negative (0)	Predicted Positive (1)
Actual Negative (0)	306	14

Figure 11 - AHI Prediction Logistic Regression Model Confusion Matrix (Train)

TEST DATA	Predicted Negative (0)	Predicted Positive (1)
Actual Negative (0)	73	7

Figure 12 - AHI Prediction Logistic Regression Model Confusion Matrix (Test)

Running the Logistic Regression model with the appropriate predictor values, we see that the majority of p-values are significant at an α level of 0.05. Analyzing the precision of our model, we see in the confusion matrices above that it is 95.625% accurate in classifying patients with severe AHI scores in the training data. When predicting on the testing data, the model had a 91.25% accuracy.

LIMITATIONS

Even though the Epworth Scale is widely regarded as a reliable measure of daytime sleepiness, there are certainly drawbacks associated with the process. Correlation values of 0.49, 0.35, and 0.52 does indicate a moderate correlation between ESS and those three variables- age, gender, and height. On top of that, there is a significant relationship between the ESS score of males and their leg movements in sleeping. However, there is a large possibility that errors can occur as a result of either misrepresentation in predictor variables or unreliability of the ESS scale.

Most obviously, ESS is a subjective yet self-reported measure, as opposed to other response variables such as BMI, AHI, or PLM. A patient, for instance, could either underestimate or overestimate their tendency to fall asleep in a certain activity. On top of that, the boundary between daytime and nighttime is considered blurry since there is no clear cutoff line. If other variables are not considered, there will be an increased variability and can potentially affect the validity of the results. In order to improve the accuracy of the data, objective measures are required by carefully monitoring the distinct sleep behaviors of the participants.

Aside from that, there could be limitations in the predictors that were selected for modeling. While a considerable amount of predictors was selected, there could still be ones that were missing. In addition, such a dataset has been subjected to data cleaning, which involves removing missing values, outliers, and other forms of errors. Despite efforts to obtain an improved version of the modeling data, such a process can very likely introduce bias or information loss. Most importantly, a logistic regression model only provides approximation associations rather than determining causal relationships between the predictors and outcome

variables. As a result, other various modeling methods, such as random assignment, may be recommended.

QUESTION #2

Does the degree of hypoxemia (low oxygen) defined using % time with an oxygen saturation less than 88 %, duration and frequency of respiratory events, correlate with burden of cardiac arrhythmia (duration and frequency of abnormal heart rhythm).

EXPLORATORY DATA ANALYSIS

The extent of our data regarding cardiac arrhythmia was text information gathered and recorded by sleep technicians within the lab. This text data is stored under the variable name SleepStudy:EKG Analysis. In order to perform our analysis, this text was converted into a binary variable, with 1 indicating the presence of noteworthy cardiac arrhythmia, and 0 otherwise. Because the text data was not presented in a standardized format, we decided to assign a coding of 1 to any observations indicating the presence of PAC, sinus arrhythmia, atrial fibrillation (or A-Fib), tachycardia, and bradycardia; all other observations received a coding of 0. Words such as “rare,” “few,” and “possible” were not considered in the coding due the the small size of the dataset, the fact that almost all non-NSR observations used such language, and the inherent subjectivity of those descriptors.

METHODOLOGY/RESULTS

Once we had our cardiac arrhythmia data in binary form, we fit a few generalized linear regression models relating arrhythmia with oxygen desaturation. A generalized linear regression model, also known as a GLM, is utilized when the response variable does not follow a normal distribution. In our case, arrhythmia follows a binomial distribution, which we can specify when training our model. Using a random sample of 80% of our data as the training data, we fit three GLM models, each with a varying level of oxygen desaturation counts as the explanatory variable (Desats < 90, Desats < 80, Desats < 70). In all three of our models, the coefficient of the Desats term was found to not be statistically significant at the $p < .05$ significance level, with p-values of .985, .835, and .579 for the Desats < 70, < 80, and < 90 models respectively. This alone is enough to conclude that the models do not have predictive value; however, we also produced the following receiver operating characteristic (ROC) curves shown below using predictions obtained when applying our models to the 20% of the data not used in the training set. The graphs plot the sensitivity (or true positive rate) against the specificity (or true negative rate). We can then use the area under the curve (AUC) to evaluate how well the model performs. Note that a random classifier should have an AUC of .5. For the models we constructed, the AUC values were .5, .5403, and .5439 for LT 70, LT 80, and LT 90 respectively. Given the proximity of these values to the .5 AUC value of a random classifier, these results further

confirm that the Desat values are not useful for predicting arrhythmia in the dataset used in this report.

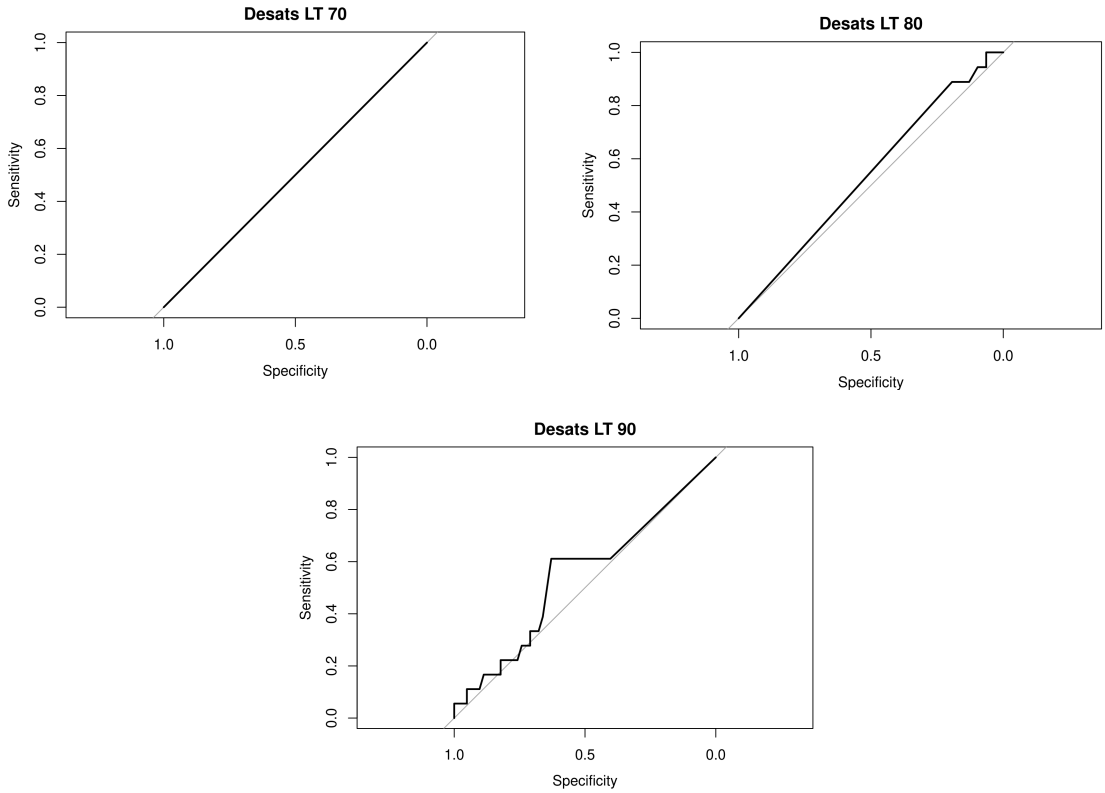

Figure 13, 14, 15 - ROC Plots for Models with Varying Desaturation Level Counts as Response

LIMITATIONS

The main limitation with our analysis was in the assigning of binary values to the text data regarding arrhythmia. While it was necessary given the lack of standardization within the text, binary values do not capture the full scope of cardiac abnormalities. Not all cardiac arrhythmia is of the same frequency, but we were forced to put any presence of abnormalities within the same category.

Additionally, the oxygen desaturation counts were given as a total, but without a variable relating to sleep time, we were unable to calculate oxygen desaturation rates. This could potentially cause issues with our analysis, because we cannot fully represent the relative frequency of oxygen desaturation events across different patients. For example, if patient A and patient B both experience 10 oxygen desaturation events during their sleep, but patient A was asleep for eight hours as opposed to four, our oxygen desaturation variable would not represent the discrepancy between the patients. For future analysis, it would be useful to have a sleep time variable, in

order to calculate desaturation rates, in order to better compare patients who may not have a consistent time asleep.

DISCUSSION

Although the extent of our data collection regarding cardiac abnormalities was obtained via text data written by technicians, we think it would be interesting to extract numerical data from EKGs to compare with our current sleep data. This could even be in the form of a standardized scale followed by all of the technicians, with 0 representing no cardiac arrhythmia, and each number between 1 and 10 representing a varying level of cardiac abnormality. This would offer a more meaningful link between discrepancies between patients cardiac rhythm and variables like oxygen desaturation.

Additionally, we think it would be valuable to investigate the impact of circadian rhythm on a patient's sleep quality. Specifically, how does the time of day or routine of sleeping in the lab influence results, as opposed to a patient's normal sleep schedule or nighttime routine. For example, if a patient usually goes to bed at 2:00 AM, but is asked to begin sleeping at 11:00 PM in the lab, we wonder if this change could cause abnormalities in sleep quality.

CONCLUSION

The correlation tests and modeling conducted with the Epworth Sleepiness Scale (ESS) scores suggest a possible insignificance of sleep indices in predicting ESS. Additionally, the question #2 analysis found no significant correlation between the degree of hypoxemia and the burden of cardiac arrhythmia, indicating that oxygen desaturation values were not useful for predicting arrhythmia in the dataset. Limitations, such as inconsistent text data and lack of standardized rates, should be considered. Nevertheless, this study sets the stage for future research by proposing investigations into the influence of circadian rhythm, stratifying by sleep phase, incorporating more cardiac event data points, and distinguishing between types of sleep apnea patients. Advancing our understanding of sleep disorders requires addressing limitations and exploring additional factors with standardized measures.

Works Cited

Asghari, Alimohamad, and Fatemeh Mohammadi. “Is Apnea-Hypopnea Index a Proper Measure for Obstructive Sleep Apnea Severity?” *Medical Journal of the Islamic Republic of Iran*, Aug. 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC3917481/.

Boyes, Jennifer, et al. “The Use of an Online Epworth Sleepiness Scale to Assess Excessive Daytime Sleepiness.” *Sleep & Breathing = Schlaf & Atmung*, May 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5399063/.

“Data and Statistics.” *Centers for Disease Control and Prevention*, 12 Sept. 2022, www.cdc.gov/sleep/data_statistics.html.

Gander, P H, et al. “The Epworth Sleepiness Scale: Influence of Age, Ethnicity, and Socioeconomic Deprivation. Epworth Sleepiness Scores of Adults in New Zealand.” *Pubmed*, Feb. 2005, pubmed.ncbi.nlm.nih.gov/16171250/.

Mokhber, Somayyeh, et al. “Comparing the Excessive Daytime Sleepiness of Obese and Non-Obese Patients.” *Iranian Red Crescent Medical Journal*, 31 May 2016, www.ncbi.nlm.nih.gov/pmc/articles/PMC5027628/.

Mokhber, Somayyeh, et al. “Comparing the Excessive Daytime Sleepiness of Obese and Non-Obese Patients.” *Iranian Red Crescent Medical Journal*, 31 May 2016, www.ncbi.nlm.nih.gov/pmc/articles/PMC5027628/.

Quan, Stuart F. “Abuse of the Epworth Sleepiness Scale.” *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*, 15 Oct. 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC3778187/.