Causal Relationship between AHI (Sleep Apnea Index) and SaO2 (Oxygen Saturation) In Sleep Health Patients

Introduction:

Causal modeling between Apnea-Hypopnea Index (AHI) and arterial oxygen saturation (SaO2) explores the intricate relationship between sleep-disordered breathing and oxygen levels in the bloodstream. AHI measures the severity of sleep apnea by tallying the number of apneas (pauses in breathing) and hypopneas (shallow breathing episodes) per hour of sleep, while SaO2 indicates the amount of oxygen carried by hemoglobin in the blood. Understanding the causal pathways between these variables can deduce how sleep disturbances affect oxygenation and vice versa, offering insights into the mechanisms underlying sleep-related health conditions like obstructive sleep apnea and their physiological impacts.

Causal Models:

- Association: This refers to the statistical relationship between two variables, indicating how changes in one variable may be related to changes in another, that is, $P(y|x)$. Here we are considering factors like age, AHI, smoking and check if these factors affect the sleep health of the patients or not.
- Cofounders: Identifying potential confounding variables that may influence both AHI and SaO2, include demographic factors (e.g., age, gender), lifestyle factors (e.g., smoking status), clinical factors (e.g., BMI), and other relevant variables.
- Intervention: To study the impact of surgery on sleep health, intervention refers to actively implementing a change, in this case, a surgical procedure, to observe its effects, i.e., $P(y|do(x), z)$. Here, $do(x)$ points to the occurrence of surgery, $do(x=1)$ means having surgery, and $do(x=0)$ means not having surgery.
- Counterfactuals: This is a hypothetical scenario that represents what would have happened in the absence of the intervention. For example, was it surgery that reduce AHI, that is, $P(y|do(x), z)$. $do(x)$ means $do(x=1)$ means having surgery, and $do(x=0)$ means not having surgery that is, $P(y_x|x',y')$.

Related Work:

We have performed causal modeling using causal inference library – 'DoWhy'.
'DoWhy' is a Python library for causal inference that provides a unified framework for estimating causal effects from observational data. It allows users to perform various causal inference tasks, including identification, estimation, and inference, using a simple and intuitive interface.

To identify and estimate the causal relationship between AHI and SaO2 the following steps are followed:

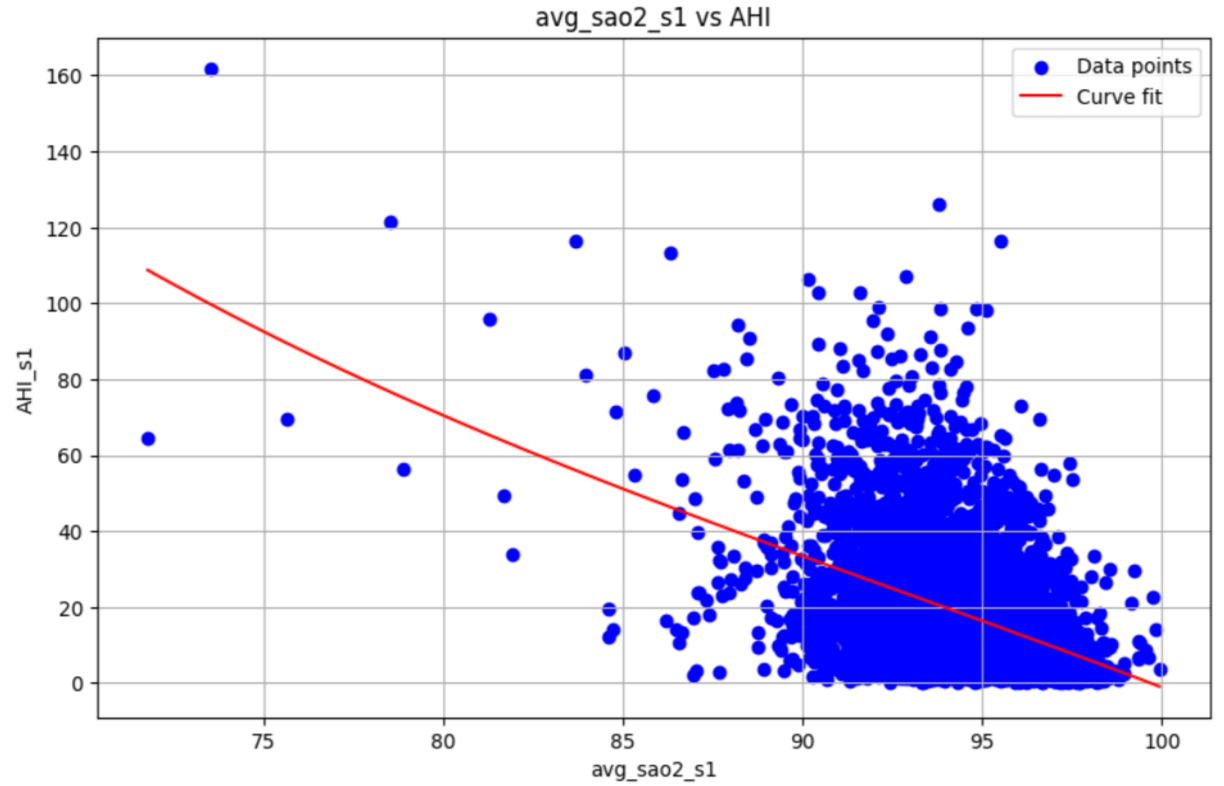
1. Data Preparation:

The Sleep Heart Health Study (SHHS) is a multi-center cohort study implemented by the National Heart Lung & Blood Institute to determine the cardiovascular and other consequences of sleep-

disordered breathing. It tests whether sleep-related breathing is associated with an increased risk of coronary heart disease, stroke, all-cause mortality, and hypertension. In all, 6,441 men and women aged 40 years and older were enrolled between November 1, 1995 and January 31, 1998 to take part in SHHS Visit 1. During exam cycle 3 (January 2001- June 2003), a second polysomnogram (SHHS Visit 2) was obtained in 3,295 of the participants. From SHHS1 and SHHS2 we have collected the sleep data (sleep stage information) patients who has undergone snoring surgeries along with their smoking habits, any other existing cardiac diseases to build a causality model to check how these impacted on the sleeping health of the patients.
The given SHH data is cleaned and preprocessed, ensuring that missing values are handled appropriately, variables are appropriately scaled and formatted, and all the potential outliers are removed.

2. Exploratory Data Analysis (EDA):

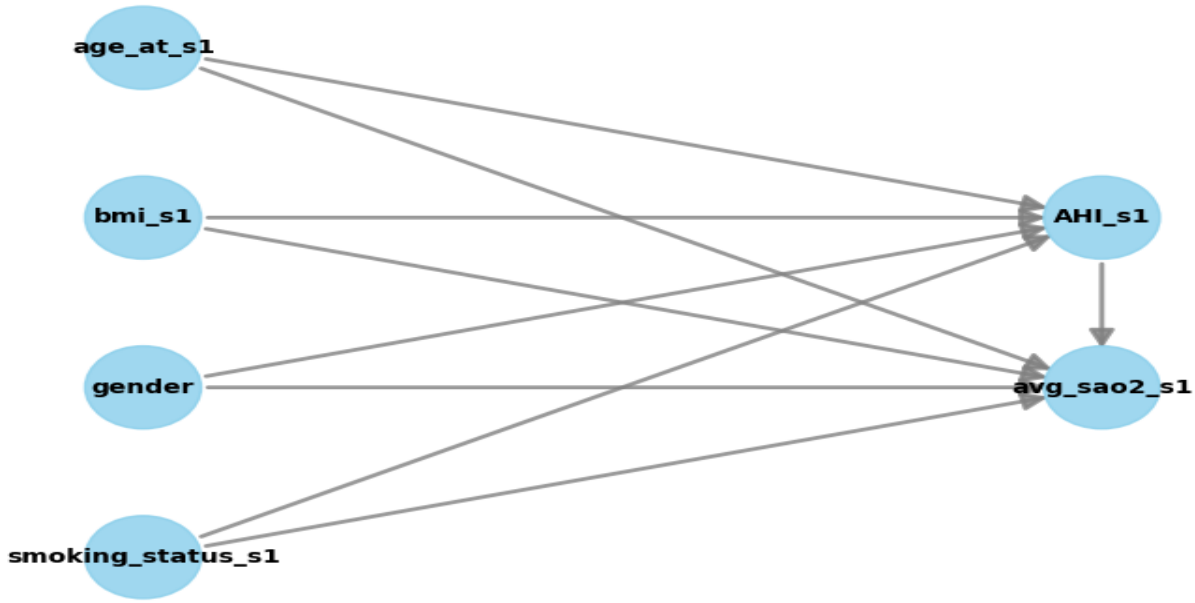
EDA is conducted to visualize the distributions of AHI and SaO2 and explore their relationship. The visualization can be deduced from the graph below:



Correlation coefficient of -0.44 indicates a moderately strong negative linear relationship between Avg SaO2 and AHI. This which means they are inversely proportional i.e. decrease in Sao2 causes increase in Ahi values and vice-versa.

3. Identifying Cofounders:

Identifying potential confounding variables that may influence both AHI and SaO2, include demographic factors (e.g., age, gender), lifestyle factors (e.g., smoking status), clinical factors (e.g., BMI), and other relevant variables.
Here, we have considered age, gender, BMI and smoking_status to be the confounding variables.



4. Causal Inferencing:

Causal Inferencing is performed using causal inference library – 'DoWhy'.

The below is the causal estimate which is achieved using 'DoWhy' library,

Identified estimand
Estimand type: EstimandType.NONPARAMETRIC_ATE

Estimand : 1
Estimand name: backdoor
Estimand expression:
  d(E[avg_sao2_s1|smoking_status_s1, gender,bmi_s1,age_at_s1])
$d[AHI_{s1}]$
Estimand assumption 1, Unconfoundedness: If U→{AHI_s1} and U→avg_sao2_s1 then P(avg_sao2_s1|AHI_s1, smoking_status_s1,gender,bmi_s1,age_at_s1,U) = P(avg_sao2_s1|AHI_s1,smoking_status_s1,gender,bmi_s1,age_at_s1)

Realized estimand
b: avg_sao2_s1~AHI_s1+smoking_status_s1+gender+bmi_s1+age_at_s1
Target units: ate

Estimate

Mean value: -0.03350371778414285

The estimate can be analyzed as below:

1. Identified Estimand:
- This section describes the identified estimand, which is the causal effect that the analysis aims to estimate.
- In this case, the identified estimand is of type "EstimandType.NONPARAMETRIC_ATE" (Average Treatment Effect, Nonparametric).
- The estimand expression represents the causal effect of AHI (Apnea-Hypopnea Index) on average SaO2 (Oxygen Saturation), conditioned on smoking status, gender, BMI (Body Mass Index), and age.

2. Estimand Assumption:
- The assumption listed (Unconfoundedness) states that if there are no unobserved confounders influencing both AHI and SaO2, then the conditional distribution of SaO2 given AHI and observed confounders is independent of AHI.

3. **Realized Estimand:**
- This section describes the realized estimand, which is the specific functional form used to estimate the causal effect.
- The realized estimand is expressed as a linear regression model, where the outcome variable is avg_sao2_s1 (average SaO2) and the predictors include AHI, smoking status, gender, BMI, and age.

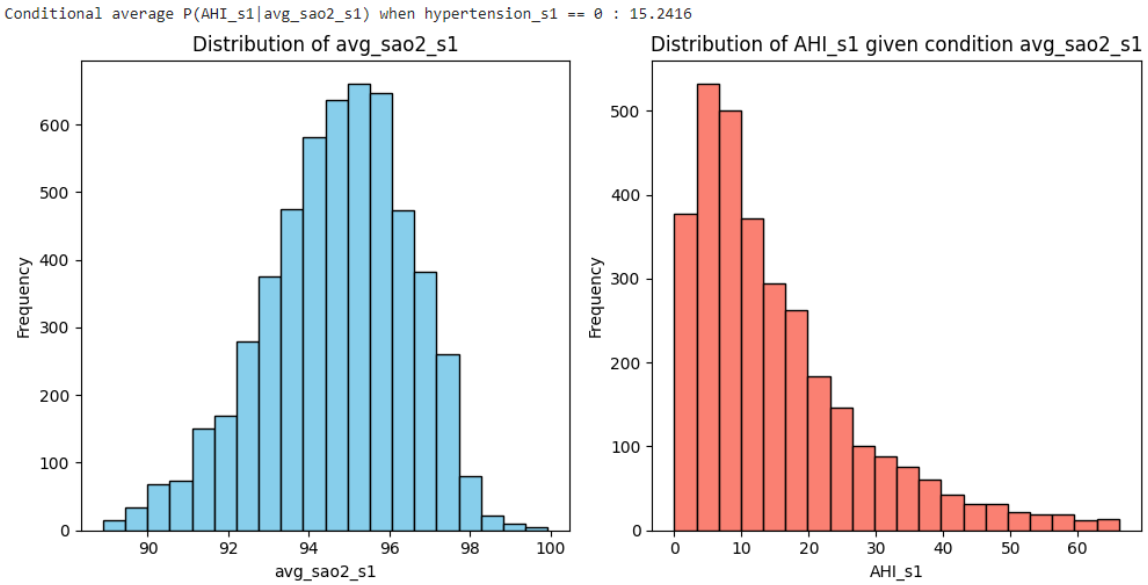
4. **Target Units:**
- The target units specify the units in which the causal effect is estimated. In this case, it's the Average Treatment Effect (ATE).

5. Estimate:
- The estimate section provides the estimated value of the causal effect.
- The mean value of the estimated causal effect is approximately -0.0335.

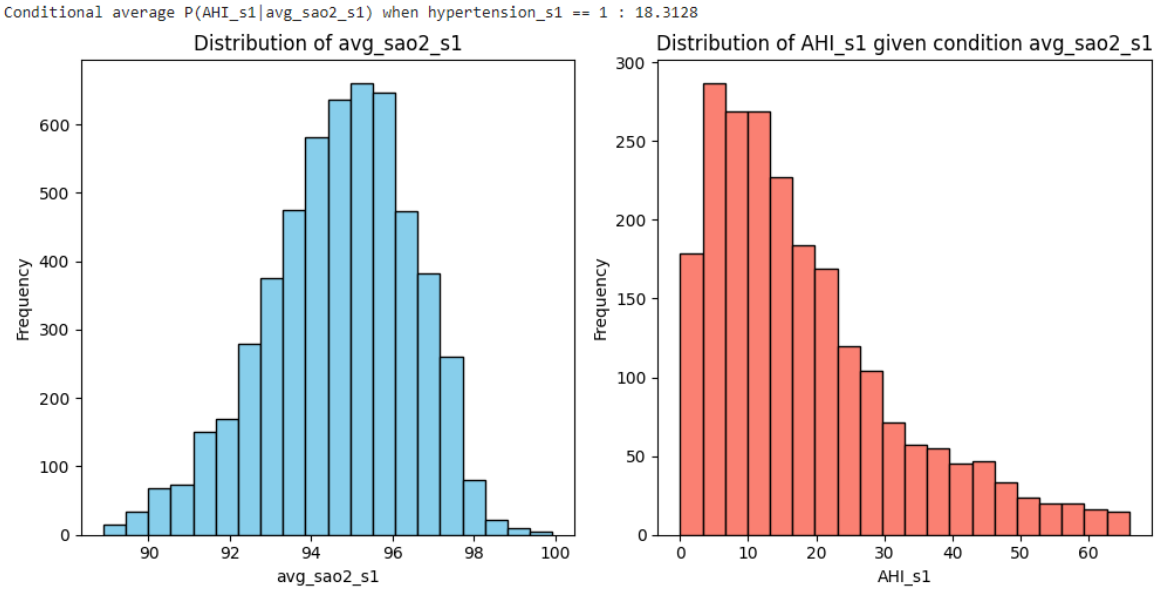
Interpreting the estimate:
- The negative sign indicates a negative relationship between AHI and average SaO2, suggesting that higher values of AHI are associated with lower average SaO2.
- The magnitude of the estimate (-0.0335) represents the average change in SaO2 (in the units of the outcome variable) for a one-unit increase in AHI, while holding other variables constant.

The above points can be interpreted by the graphs below:

When Hypertension = 0:

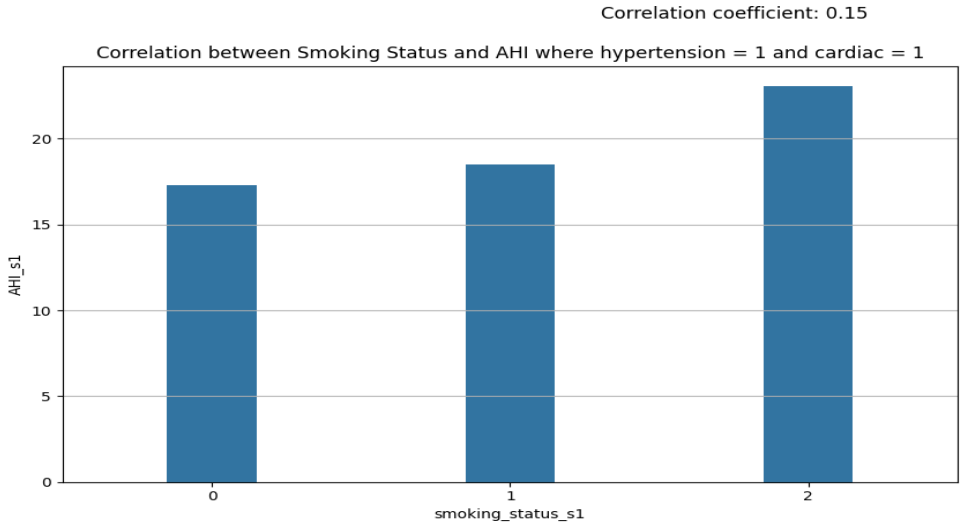


When Hypertension = 1:



In both the distributions above AHI and SaO2 almost remains same, irrespective of hypertension.

- But we can see that smoking does effect AHI from the below graph, as previous smokers have more AHI values compared to non-smokers or the patients who have just stated smoking.



0.0 – never smoked
1.0 - currently smoking
2.0 – was a previous smoker

Patients who have quit smoking (was previous smokers), have comparatively more AHI values, concluding there may be some residual effects of smoking even after quitting, effecting their respiratory function, and contributing to higher AHI.

Conclusion:

From the above interpretations, we can conclude that, there is a causal inference between AHI and SaO2, as we can observe conditional probability of P(avg_sao2_s1|AHI_s1) considering smoking_status of the patients might be the cause for higher AHI values.
As SaO2 levels decreases, AHI levels increase, due to the impact of sleep apnea on respiratory functions. Elevated AHI leads to severity of breathing disturbances during sleep. The ideal level of oxygen is usually 95% or higher. We can see that, the patients who have higher AHI values have lower SaO2 levels (less than 95%).

References:

- https://learning.oreilly.com/library/view/causal-inference-and/9781804612989/B18993_Part_1.xhtml
- **https://mixtape.scunning.com/**
- **https://www.pywhy.org/dowhy/v0.11.1/**
- **https://sleepdata.org/**