Module_1:

Team Members: Aarnav Shah & Shreesh Kalagi

Project Title: Atherosclerosis and Beta Amyloid Accumulation

Project Goal:

Research Question: Does atherosclerosis correlate with amyloid beta accumulation and lead to higher THAL scores in patients with Alzheimer's disease?

By using the THAL score as a way to quantify beta amyloid presence in the brain, the goal of the investigation is to investigate the relationship between the condition of atherosclerosis and beta amyloid accumulation in Alzheimer's Patients.

Disease Background: Alzheimer's

- Prevalence & incidence
  - Prevalence: 6.9 million / 340.1 million x 100 = 0.0203 x 100 = 2.03% in the USA (2024)
  - Incidence: The number could double to 13.8 million by 2060 (not factoring improvements in treatment or possible cures)
- Economic burden: In 2020, the estimated economic burden was estimated at 305 billion USD
  - This likely underestimates indirect costs, like the productivity lost through caretaking by family members
- Risk factors (genetic, lifestyle)
  - Age, especially after 65
  - Genetics (family history, genes such as APOE4)
    - Other genes such as PSEN1/2 and APP can cause early onset Alzheimer's
  - Gender: Women get Alzheimer's more often than men
  - Smoking
  - Excessive consumption of alcohol
  - Sedentary lifestyle
  - Stress
  - High blood pressure
  - Occurence of traumatic brain injury

    - Sleep issues
- Societal determinants
    - Level of education: People without a high-school diploma present with the symptoms of Alzheimer's more frequently and earlier in their lvies
    - Because other medical conditions can lead to Alzheimer's risk, poor access to healthcare increases the incidence of Alzheimer's
    - Social isolation has been linked to dementia
- Symptoms
    - Declining memory (especially with recent events and learning new things)
    - Declining cognition (harder to concentrate, plan, etc.)
    - Irritability and restlessness
    - Delusions
    - Inability to maintain lifestyle (in terms of eating, sleeping)
- Diagnosis
    - Neurological Exam (reflexes, coordination, and speech are tested)
    - Cognitive / behavioural tests (AD8, SLUMS, etc. These may also be digital)
    - Depression Screening (as it's one of the symptoms)
    - Genetic Testing (for APOE4 etc.)
    - Brain imaging (eliminates other conditions)
    - Presence of beta amyloid in the blood
- Standard of care treatments (& reimbursement)
    - Early stage Alzheimer's
        - Galantamine, benzgalantamine, rivastigmine, and donepezil can treat some of the cognitive and behavioural symptoms if they're mild
            - Covered in medicare part D
        - Cholinesterase inhibitors prevent the breakdown of Acetylcholine (a chemical important for thinking), but since the issue is the production of the chemical and not the breakdown, this isn't effective for too long
        - Lecanemab and donanemab target beta amyloid and reduce plaque formation (these are the main treatments, they don't just suppress symptoms)
            - Insurance (Medicare Part B) may only cover this in certain situations
    - Mid to late stage Alzheimer's
        - Memantine and Donepzil regulates glutamate, preventing its overproduction which would cause cell death. It helps maintain normal function
            - Covered in medicare part D
- Disease progression & prognosis
    - According to Mayo Clininc, Alzheimer's Disease progesses gradually over years in stages
    - Preclinical Alzheimer's disease
    - Mild cognitive impairment due to Alzheimer's disease
    - Mild dementia due to Alzheimer's disease

    - Moderate dementia due to Alzheimer's disease
    - Severe dementia due to Alzheimer's disease
- Continuum of care providers
    - Physician care: Primary Care, Neurologists, Geriatricians
    - Non-physician care: Patient Care Assistants, Medication Aides, Home nurses, Hospice care
- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology)
    - Alzheimer's Disease (AD) causes neurodegeneration through the destruction of inter-neural links in the brain. Extracellular accumulation of amyloid-Beta plaques causes disruption in synapses, while entanglement of excess Tau proteins intracellularly contributes to neuronal death.
    - AD causes progressive cognitive decline by damaging neural activity in several parts of the brain, including the entorhinal cortex, hippocampus, cortex, thalamus, hypothalamus, locus ceruleus, and amygdala.
- Clinical Trials/next-gen therapies
    - NIH funded AD clinical trials are separated by pharmacological and non-pharmacological studies...
        - Pharmacological: Includes studies of inflammation, vasculature, metabolism, circadian rhythm, and relevant protein activity (amyloids, tau, etc.)
        - Non-pharmacological: Includes modality and lifestyle factor, such as exercise, neurostimulation, cognitive training, sleep-related studies, etc.

Prevalence and incidence: https://pubmed.ncbi.nlm.nih.gov/38689398/
Economic Burden: https://pubmed.ncbi.nlm.nih.gov/32840331/
Risk factors: Google Gemini
Societal determinants: https://www.cdc.gov/alzheimers-dementia/php/sdoh/index.html
Symptoms: https://www.alzheimers.org.uk/about-dementia/types-dementia/alzheimers-disease-symptoms
Standard of care: https://www.nia.nih.gov/health/alzheimers-treatment/how-alzheimers-disease-treated and ChatGPT (for the insurance information)
Biological Mechanisms: https://pmc.ncbi.nlm.nih.gov/articles/PMC5290713/
Disease Prognosis: Mayo Clinic Clinical trials: https://www.nia.nih.gov/research/ongoing-AD-trials

Data-Set:

The dataset being used is from the paper titled *Integrated multimodal cell atlas of Alzheimer's disease*, published in the Nature Neuroscience journal. It's open access, which allows us to use it for the assignment as long as we give credit to it.

It includes data from 33 male and 51 female donors, for a total of 84 people, all of which are usable for our investigation. The metrics incldued are extensive, including lifestyle factors like

years of education, behavioural and cognitive testing scores like MMSE, biological information like beta amyloid presence, etc.

For our investigation 'Does atherosclerosis correlate with amyloid beta accumulation and lead to higher THAL scores in patients with Alzheimer's disease?' we used the following data: Thal Score (0-5) Arteriosclerosis (None, Mild, Moderate, Severe)

They are both catergorical data.

Gabitto, M.I., Travaglini, K.J., Rachleff, V.M. et al. Integrated multimodal cell atlas of Alzheimer's disease. Nat Neurosci 27, 2366–2383 (2024). https://doi.org/10.1038/s41593-024-01774-5

Data Analyis:

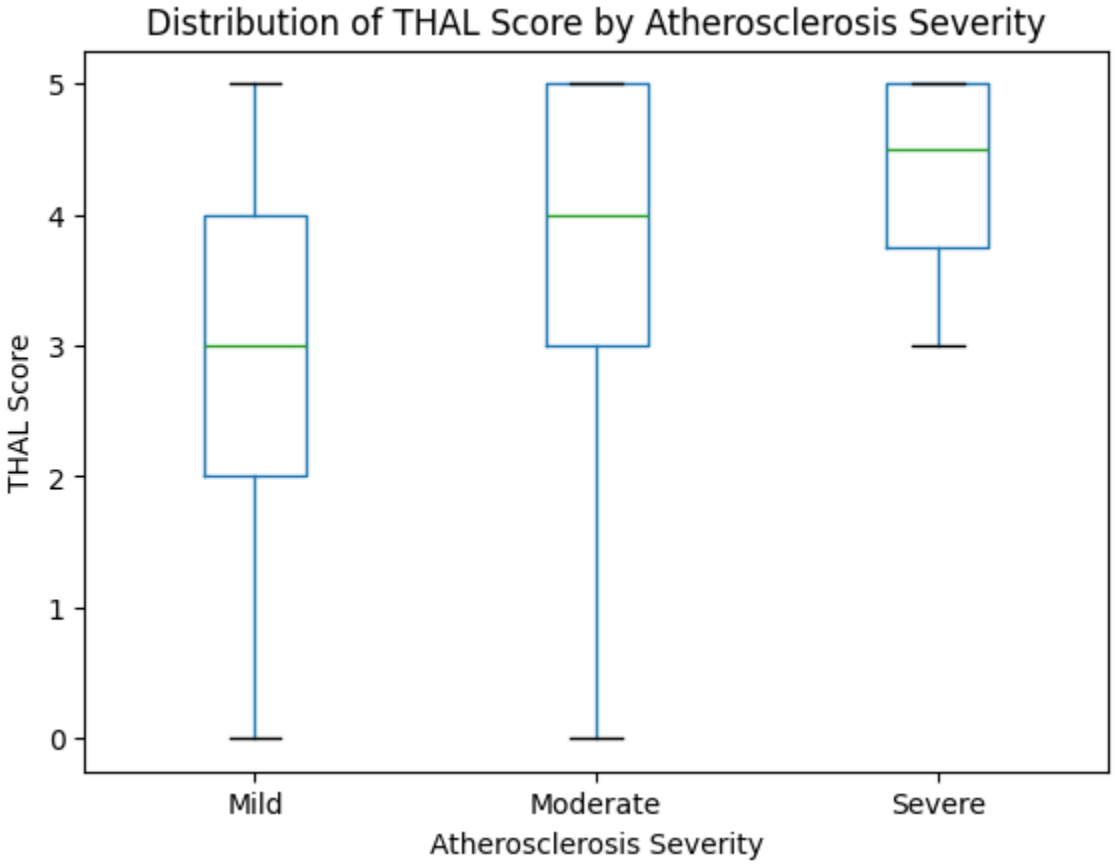
In [1]:
```
# The Patient class
import pandas as pd

class Patient:
        all_patients = []

        def __init__(self, donor_id, atherosclerosis, THAL, ABeta40, ABeta42, CASI)
                self.donor_id = donor_id
                self.atherosclerosis = atherosclerosis
                self.THAL = THAL
                self.ABeta40 = ABeta40
                self.ABeta42 = ABeta42
                self.CASI = CASI
                Patient.all_patients.append(self)

        def __repr__(self):
                return f"Patient(ID: {self.donor_id}, Atherosclerosis: {self.athero

        def get_id(self):
                return self.donor_id
        def get_atherosclerosis(self):
                return self.atherosclerosis
        def get_THAL(self):
                return self.THAL
        def get_ABeta40(self):
                return self.ABeta40
        def get_ABeta42(self):
                return self.ABeta42
        def get_CASI(self):
                return self.CASI

        @classmethod
        def combine_and_instantiate(cls):
                with open("data/UpdatedLuminex.csv") as f:
                        luminex = pd.read_csv(f)

                with open("data/UpdatedMetaData.csv") as f:
```

```
                metadata = pd.read_csv(f)

        # merge on id
        merged_df = pd.merge(luminex, metadata, on='Donor ID')

        for index, row in merged_df.iterrows():
                cls(
                        donor_id = row['Donor ID'],
                        atherosclerosis = row['Atherosclerosis'],
                        THAL = row['Thal'],
                        ABeta40 = row['ABeta40 pg/ug'],
                        ABeta42 = row['ABeta42 pg/ug'],
                        CASI = row['Last CASI Score']
                )

        cls.all_patients.sort(key = cls.get_id)
```

In [2]:
```
# main.py - Analysis code

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re

# Calls the class method to combine data
Patient.combine_and_instantiate()

data = []

# Creates a list of dictionaries to make a DataFrame
for patient in Patient.all_patients:
    data.append({
        'Donor ID': patient.get_id(),
        'Atherosclerosis': patient.get_atherosclerosis(),
        'THAL': patient.get_THAL(),
        'ABeta40': patient.get_ABeta40(),
        'ABeta42': patient.get_ABeta42(),
        'CASI': patient.get_CASI()
    })

# Pandas dataframe creation
df = pd.DataFrame(data, columns=['Donor ID', 'Atherosclerosis', 'THAL', 'THAL_score

# Convert severity levels to numeric (For ordering)
severity_order = ["mild", "moderate", "severe"]

# Extracts the numerical part of the THAL string (it's formatted like "Thal X" in t
df["THAL_score"] = df["THAL"].apply(
    lambda x: int(re.search(r'\d+', str(x)).group())
)
```

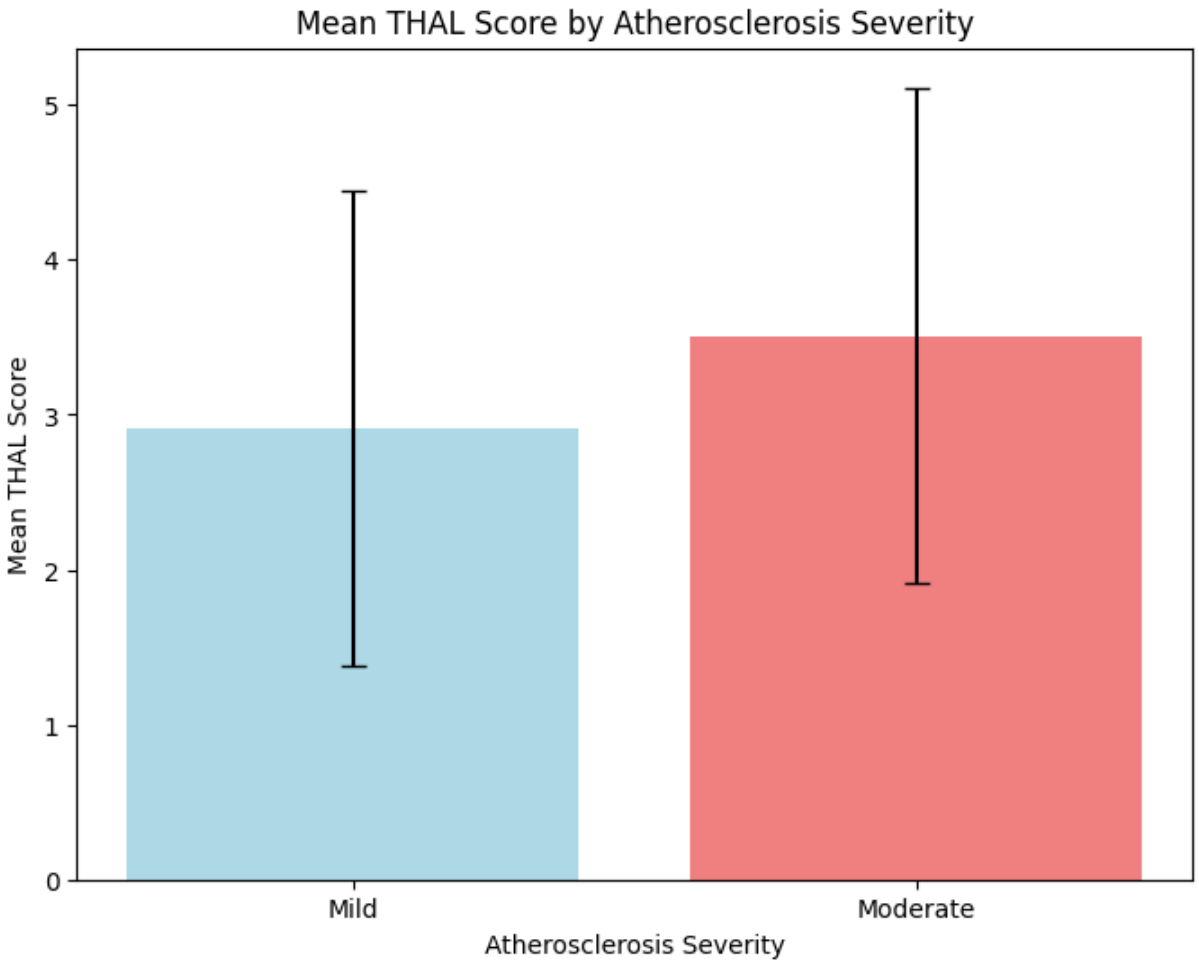
In [3]:
```
# Boxplot of THAL score by Atherosclerosis severity
df.boxplot(column="THAL_score", by="Atherosclerosis", grid=False, positions=range(l

plt.xlabel("Atherosclerosis Severity")
```

```
plt.ylabel("THAL Score")
plt.title("Distribution of THAL Score by Atherosclerosis Severity")
plt.suptitle("")  # removes the default pandas boxplot title
plt.show()
```



In [4]:
```
# T-test between mild and moderate atherosclerosis groups (they have more samples)
# One sided because amyloid beta is expected to increase with atherosclerosis
# Unpaired because the patients aren't the same
from scipy import stats
mild_scores = df[df["Atherosclerosis"] == "Mild"]["THAL_score"].dropna()
moderate_scores = df[df["Atherosclerosis"] == "Moderate"]["THAL_score"].dropna()

# Are the variances equal?
print("mild score variance", np.var(mild_scores), " moderate score variance", np.va

# Bar Graph
plt.figure(figsize=(8, 6))
# Error bars represent standard deviation
plt.bar(['Mild', 'Moderate'], [np.mean(mild_scores), np.mean(moderate_scores)], yer
plt.ylabel('Mean THAL Score')
plt.xlabel('Atherosclerosis Severity')
plt.title('Mean THAL Score by Atherosclerosis Severity')
plt.show()

t_statistic, p_value = stats.ttest_ind(mild_scores, moderate_scores, equal_var=Fals
print("T-test results:")
print(f"T-statistic: {t_statistic}, P-value: {p_value}")
```

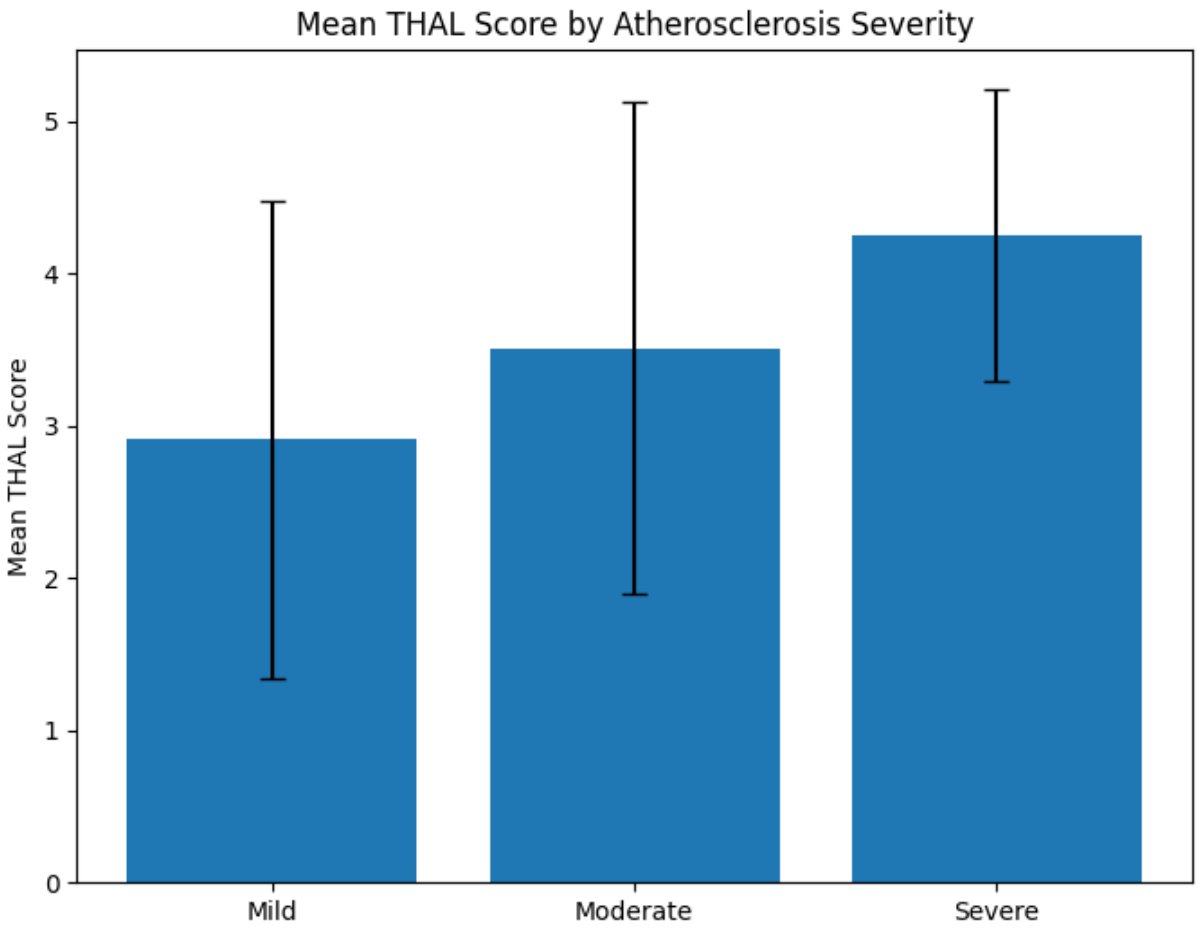
```
mild score variance 2.355371900826446  moderate score variance 2.5636293733179554
```




```
T-test results:
T-statistic: -1.4859252354956645, P-value: 0.1449502578577202
```

Statistically insignificant, since our p-value 0.14 > 0.05

In [5]:
```
# ANOVA Test
group1 = df[df["Atherosclerosis"] == "Mild"]["THAL_score"].dropna()
group2 = df[df["Atherosclerosis"] == "Moderate"]["THAL_score"].dropna()
group3 = df[df["Atherosclerosis"] == "Severe"]["THAL_score"].dropna()

# Bar plot
plt.figure(figsize=(8, 6))
plt.bar(['Mild', 'Moderate', 'Severe'], [group1.mean(), group2.mean(), group3.mean(
plt.ylabel('Mean THAL Score')
plt.title('Mean THAL Score by Atherosclerosis Severity')
plt.show()

# One way ANOVA to compare means across the three groups
f_statistic, p_value = stats.f_oneway(group1, group2, group3)
print("ANOVA results:")
print(f"F-statistic: {f_statistic}, P-value: {p_value}")
```

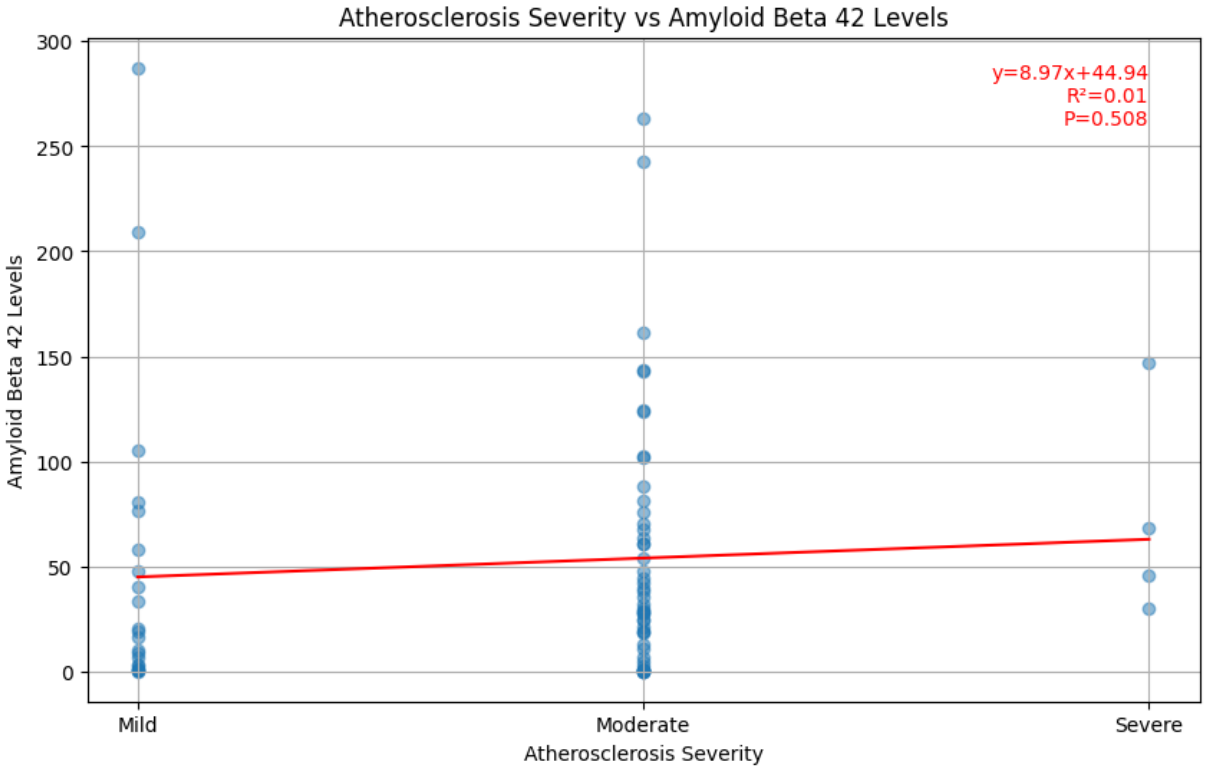



```
ANOVA results:
F-statistic: 1.7499650451870399, P-value: 0.18089440413763147
```

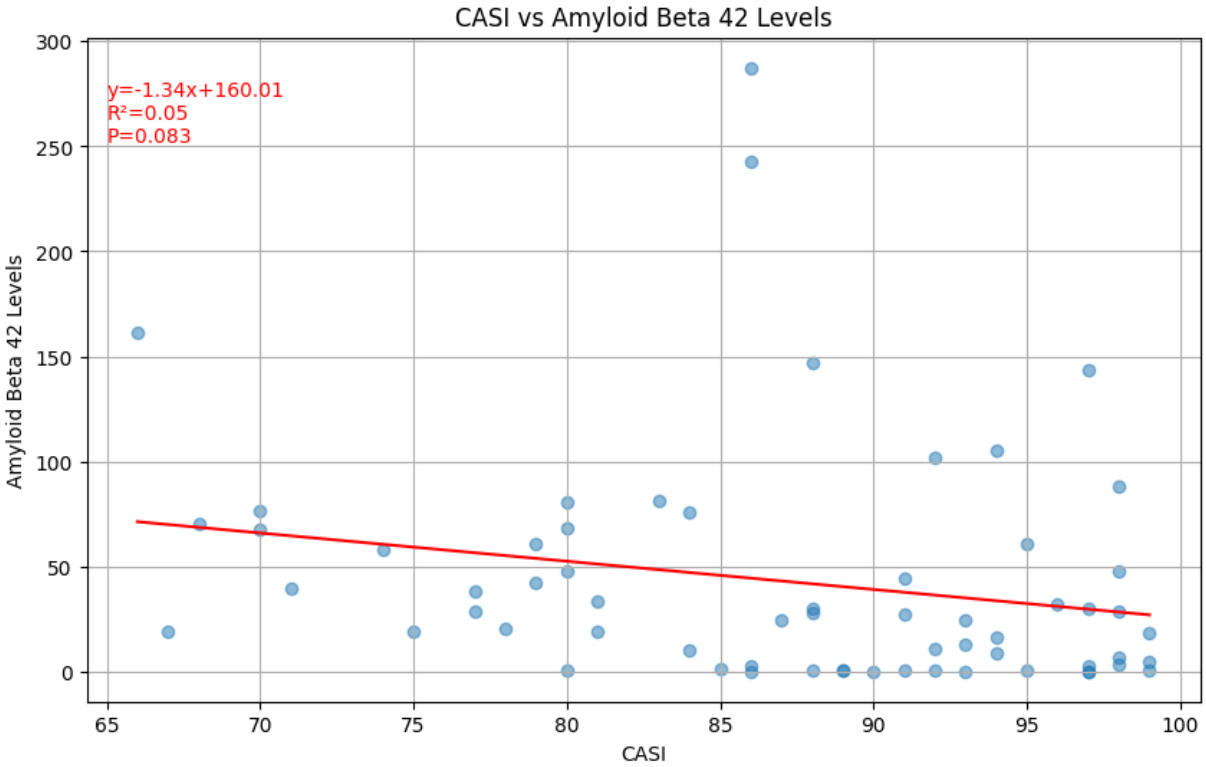
Statistically insignificant, as our p-value of 0.18 > 0.05

In [ ]:
```
# Scatter Plot of Atherosclerosis Severity vs Amyloid Beta Levels

plt.figure(figsize=(10, 6))

# Filtering out the outlier ABeta42 value that's greater than 1000
df = df.dropna(subset=["Atherosclerosis", "ABeta42"])
df = df[df["ABeta42"] < 1000]

# Scatter plot
plt.scatter(df["Atherosclerosis"], df["ABeta42"], alpha=0.5) # Alpha determines opa


# Linear Regression
severity_numeric = df["Atherosclerosis"].map({"Mild": 0, "Moderate": 1, "Severe": 2
slope, intercept, r_value, p_value, std_err = stats.linregress(severity_numeric, df
x = np.array([0, 1, 2])
y = slope * x + intercept
r_squared = r_value**2 #type: ignore

plt.text(2, 290, f'y={slope:.2f}x+{intercept:.2f}\nR²={r_squared:.2f}\nP={p_value:.

plt.title('Atherosclerosis Severity vs Amyloid Beta 42 Levels')
plt.xlabel('Atherosclerosis Severity')
```

```
plt.ylabel('Amyloid Beta 42 Levels')

plt.plot(x, y, color='red', label='Regression Line')

plt.grid()
plt.show()
```



In [7]:
```
# This plot is just for using a scatter plot with two continunous variables
# ABeta42 vs CASI
plt.figure(figsize=(10, 6))
df = df.dropna(subset=["CASI", "ABeta42"])
df = df[df["ABeta42"] < 1000]
plt.scatter(df["CASI"], df["ABeta42"], alpha=0.5)
# Linear Regression
slope, intercept, r_value, p_value, std_err = stats.linregress(df["CASI"], df["ABet
x = np.array([df["CASI"].min(), df["CASI"].max()])
y = slope * x + intercept
r_squared = r_value**2 #type: ignore

plt.text(65, 250, f'y={slope:.2f}x+{intercept:.2f}\nR²={r_squared:.2f}\nP={p_value:

plt.title('CASI vs Amyloid Beta 42 Levels')
plt.xlabel('CASI')
plt.ylabel('Amyloid Beta 42 Levels')

plt.plot(x, y, color='red', label='Regression Line')


plt.grid()
plt.show()
```



In [8]:
```
# Export to csv
df.head()
df.to_csv("combined_patient_data.csv", index=False)
```

Verify and validate your analysis:

Analysis

To investigate the relationship between Atherosclerosis severity and Amyloid Beta Accumulation, we used to the categorical data of atherosclerosis (mild, moderate, severe) and the THAL score, which quantifies amyloid beta accumulation in the brain (1-5). Our alternative hypothesis was that the degree of atherosclerosis would positively predict THAL score, which is to say a higher degree of atherosclerosis would lead to a higher THAL score.

We investigated this using a T-test (with only mild and moderate categories since they have more data points), an ANOVA test and also tried a scatter plot with Amyloid-Beta 42 as the y-axis instead of the THAL score. The p-values were 0.14, 0.18, and 0.51 respectively.

In all of these tests and plots, although the trend was in the direction we predicted, we had p-values greater than 0.05, which indicates the findings are statistically insignificant.

Related Literature

A paper published by Dietmar Rudolf Thal, titled, "Vascular pathology in Alzheimer disease: correlation of cerebral amyloid angiopathy and arteriosclerosis/lipohyalinosis with cognitive decline," studied the correlation of both cerebral amyloid angiopathy (CAA) and

arteriosclerosis/lipohyalinosis (AS/LH) with cognitive decline in patients with Alzheimer's Disease (AD). The paper describes CAA as arterial disease caused by deposition of amyloid proteins, with AS/LH being defined as the stiffening of arteries and build-up lipids in cerebral arteries. The results of the study indicate that "AD cases showed a higher number of regions with CAA and AS/LH compared to nondemented patients with AD-related pathology and controls." Essentially, increased presence of amyloid protein deposits and arteriosclerosis in AD patients correlates with declined cognitive ability, which is rated by the THAL Score metric in our data analysis.

However, our studies primarily focused on atherosclerosis rather than arteriosclerosis. Although atherosclerosis is a form of arteriosclerosis (both involve the stiffening of arteries, although atherosclerosis refers to stiffening specifically due to plaque build-up), we decided to search for another study that primarily focuses on the relationship between atherosclerosis and AD. To this end, a paper published in BMC Geriatrics titled, "Atherosclerosis and Alzheimer - diseases with a common cause? Inflammation, oxysterols, vasculature," similarly finds a correlation between the incidence of atherosclerosis and AD, even correlating the pathology of both conditions remarking, "We postulate that AD and ATH are both caused by chronic immunologic challenge that induces CH25H expression and protection against particular infectious agents, but at the expense of longer-term pathology." Clearly, literature marks a correlation with amyloid protein deposition, atherosclerosis, and declined cognitive ability (as represented by increased THAL Score in our data set).

The insignificance of our statistical tests and subsequent acceptance of the null hypothesis contradicts our findings in related literature. Thus, it is important to note that the complex pathology of atherosclerosis, as well as the difference between amyloid-beta protein deposition, as generalized in the above article, and global A-Beta 42, which was the metric we used in our study. Further studies must be conducted in order to either confirm or deny the correlation between our studied variables in order to determine if our question may be answered by existing knowledge in the field.

Links to papers used: https://pubmed.ncbi.nlm.nih.gov/14692704/

https://bmcgeriatr.biomedcentral.com/articles/10.1186/1471-2318-14-36

Conclusions and Ethical Implications:

We conclude that, since our statistical tests returned p-values indicative of insignificance, that there is not necessarily a correlation between amyloid beta concentrations, atherosclerosis, and THAL score in the dataset we used. However, we, as researchers and scientists, are ethically obligated to test the validity of this conclusion through further study. By simply drawing a close to the study of correlation between amyloid beta concentrations, atherosclerosis, and THAL score with the insignificance of this study alone, we fail to consider the several factors of error that may have been presented in this study. Rather, the ethically

righteous path to take from this point would be to conduct further study in this field, whether through analysis of a different, larger dataset, or through a more technical look into the mechanisms involved with amyloid beta protein deposition and the development of atherosclerosis with a wet-lab study. In any case, denying the overall possibility of significance in the question we posed would be unethical without further data to support our claim of insignificance.

Furthermore, we have an ethical obligation as researchers to use our findings in order to inform clinical care. If further studies do, in fact, confirm noncorrelation between A-Beta 42, Atherosclerosis, and cognitive decline, it is an ethical obligation that this information be relayed to clinicians such that they do not rely on the A-Beta 42 metrics in determining the atherosclerosis status of a patient, or whether both of these combined factors necessary imply a cognitive decline in Alzheimer's Disease patients. Similarly, physicians must be informed in order to ethically comply with insurance companies and others who create financial policy in the medical field, since all parties involved must be aware of this noncorrelation in order to more accurately order testing that will affect the patients' medical bills.

Limitations and Future Work:

This investigation was limited by the size of the dataset (84 patients), which could explain why the tests yielded no statistical significance. The outlier with an abnormally high beta amyloid levels may also suggest poor quality of data. However, in terms of experimental design, we were limited by using two categorical variables, which made tools like scatter plots less effective. Using raw biomarker data instead of scores or indexes like the THAL and Atherosclerosis classification might have given more conclusive results.

This is a relatively new dataset, using a new luminex assay techinque. As more data is collected and larger datasets are built, the dataset size may lead to stronger, statistically significant conclusions. So, this may be a research question worth revisiting down the line, considering that the trend was consistent with the prediction.