Module_1:

Team Members:

Hudson King and Leah Devendorf

Project Title:

Biomarker Ratio's Role in Alzheimer's Disease

Project Goal:

- This project seeks to determine the effectiveness in determining a patient's overall Alzehmer's Disease diagnoses via measuring their the ratio between their pTau and amyloid beta-42 biomarker levels.
- Is the measurement of pTau:AB42 ratio levels significant enough to lead to an AD diagnoses?
- We hypothesize that an increase in pTau:AB42 ratio levels will result in an increase of AD symptoms across numerous demographics.

Disease Background:

(Leah) I used the following pdf from 2025 as the main source. Alzheimer's Association 2025 Alzheimer's Disease Facts and Figures

- Prevalence & incidence
  - It is estimated that 7.2 million Americans are living with Alzheimer's in 2025; roughly 1 in 9 people over the age of 65. https://www.alz.org/getmedia/ef8f48f9-ad36-48ea-87f9-b74034635c1e/alzheimers-facts-and-figures.pdf
- Economic burden
  - In 2020, an estimated $305 billion was spent on taking care of people with Alzheimer's. https://pubmed.ncbi.nlm.nih.gov/32840331/
  - Most of the aid given to those suffering from Alzheimer's comes from unpaid caregivers. There are nearly 12 million unpaid caregivers, doing an estimated 413 billion dollars' worth of labor. https://www.alz.org/getmedia/ef8f48f9-ad36-48ea-87f9-b74034635c1e/alzheimers-facts-and-figures.pdf
- Risk factors (genetic, lifestyle)
  - People with a close relative who developed Alzheimer's and people with down syndrome are more at risk for the disease.
  - People who carry the e4 form of the APOE gene are more likely to develop AD. The e2 form is known to decrease the risk of AD.
  - People with poor cardiovascular health or hypertension, are smokers, or have experienced sensory loss or a traumatic brain injury are also more at risk. https://www.alz.org/getmedia/ef8f48f9-ad36-48ea-87f9-b74034635c1e/

alzheimers-facts-and-figures.pdf https://alzheimersprevention.org/alzheimers-info/risk-factors/

- Societal determinants
  - Being of a low socioeconomic status indicates a higher likelihood of many lifestyle risk factors for Alzheimer's, such as having lower cardiovascular health, diabetes, obesity, or being exposed to more air pollution.
  - Higher education and frequent social interaction has been known to alleviate the symptoms of Alzheimer's disease.
  - Those without access to proper healthcare will have a harder time getting treatment for AD. https://www.alz.org/getmedia/ef8f48f9-ad36-48ea-87f9-b74034635c1e/alzheimers-facts-and-figures.pdf https://www.cdc.gov/alzheimers-dementia/php/sdoh/index.html
- Symptoms
  - Memory loss is the key symptom. This starts with small things like recent conversations or appointments. Over time it can evolve into forgetting family members or everyday locations.
  - Additional symptoms are poor thinking, decision making, and reasoning, especially where language and numbers are involved, along with repeating simple, routine tasks.
  - As AD progresses, they may also experience changes in personality, habits, or behavior. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447

Hudson King: (ChatGPT Provided Cited Information from Published Medical Journals using the Following Prompt: "Help me learn about Alzheimer's as a disease by giving bullet pointed sentences with answers cited published medical journals on these points:")

- Diagnosis
  - Biological process that begins with the appearance of AD neuropathic change while people are asymptomatic.
  - Progression follows with PET, Cerbrospinal Fluid (CSF) biomarkers, and plasma biomarkers mapping onto either amyloid beta (40 or 42) or AD tauopathy pathway (t(total)Tau or p(active)Tau).
  - An abnormal Core 1 biomarker result is sufficient to establish a diagnosis of Ad and inform clinical decision.
  - https://pubmed.ncbi.nlm.nih.gov/38934362/
- Standard of care treatments (& reimbursement)
  - ChEIs, associated with cognitive benefits, reduce mortality risk over time; Galantamine, a ChEI, demonstrates significant reduction in developing severe dementia.
    - https://www.neurology.org/doi/10.1212/WNL.0000000000011832?
  - ChEIs increase the levels of neurotransmitter acetylcholine (ACh) in the brain, increasing memory and thinking while decreasing dementia symptoms.
    - https://pubmed.ncbi.nlm.nih.gov/36096687/#:~:text=Discussion:%20There%20is%20moderate%2Dquality,(June%2011%2C%202021).

    - Lecanemab reduced markers of amyloid in early Alzeheimer's disease and resulte in moderately less decline on measure of cognition, but it was associated with adverse events.
        - https://www.nejm.org/doi/full/10.1056/NEJMoa2212948?
- Disease progression & prognosis
    - Average survival after Alzheimer's dementia diagnosis is greater than or equal to 65 yearsm with and average range of 4-8 years and wide variability depending on age, comorbidity, and sex.
        - https://pmc.ncbi.nlm.nih.gov/articles/PMC11095490/?
    - Predictors of worse survival: Older Age, Male Sex, Lower baseline cognition function, Neuropsychiatric symptoms, and Comorbidities.
        - https://www.nature.com/articles/s41398-024-02897-w?
- Continuum of care providers
    - Integrated models of long term care providers: primary care, neurology, geriatrics, neuropsychology, nursing, social work, OT/PT, and community partners.
        - https://ijic.org/articles/10.5334/ijic.5675?
    - Research shows similar patient outcomes from health-system and community-based programs as a continuum of care, highlighting the importance of quality over setting.
        - https://jamanetwork.com/journals/jama/fullarticle/2829720?
- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology)
    - Extra cellular amyloid beta proteins and intracellular tao proteins disrupt electric signal sent by neurons through tanglement and blocking (clocking)
    - AB proteins can be AB-40 and AB-42; the biomarker of AB-42 is a hallmark in Alzheimer Diagnosis, as it's ability to block synapse responses are greater with the two extra strands.
    - Neuroinflammation interact with both proteins, contributing to progression of Alzheimer's Disease
        - https://jneuroinflammation.biomedcentral.com/articles/10.1186/s12974-023-02853-3?
- Clinical Trials/next-gen therapies
    - Anti-amyloid antibodies have Phase 3 evidence of modest slowing of decline; numerous anti-tau, neuroinflammation, and synaptic agents are in Phase 2/3
        - https://www.nejm.org/doi/full/10.1056/NEJMoa2212948?
    - 2024 landscape counted 164 trials testing 127 drugs across diverse targets
        - https://pubmed.ncbi.nlm.nih.gov/38659717/

Data-Set:

- Data is split into two separate tables: luminex and meta. The luminex data set corresponds to the amyloid-beta and tau level counts, with each accounting for their variants; on the other hand, the meta data set corresponds to the demographic data of each donor. Data collection occurred through numerous inputs: Whole slide imaging, Fluoresence activated nuclei sorting, Quantification of cDNA library fragment sizes, DNA library sequencing, and Obtaining spatial transcriptomic data. As for the study, donors

were obtained from the Adult Changes in Thought (ACT) Study and the University of Washington Alzeihmer's Disease Research Center (ADRC) where:
    - Human brain tissue was collected via rapid autopsy (at most 12 hours after death)
    - Informed consent was applied for all donors.
    - All measurements taken are taken from an exact sample size with a specific unit of measure.
        - https://canvas.its.virginia.edu/courses/153653/files/15906851?wrap=1
- Questions:
    - How do the ratio between amyloid beta and tau concentrations correlate to the diagnosis of Alzehmers?
    - How does the continuum of education effect the diagnosis of AD?
    - How does age correlate to the amoutns of tau and beta amyloid?
    - Does APOE-4 increase amyloid-beta and tau concentrations?

Importing and Filtering Data

```
import csv
import warnings
import matplotlib.pyplot as plt
!pip install scikit-posthocs

# Creates a patient oriented object for dat access
class Patient:

    all_patients = []

    death_age = []

    def __init__(self, DonorID, ABeta40: float , ABeta42: float, tTau:
float, pTau: float):

        self.DonorID = DonorID
        self.ABeta40 = ABeta40
        self.ABeta42 = ABeta42
        self.tTau = tTau
        self.pTau = pTau
        self.apoe_genotype = None
        self.death_age = None
        self.cog_stat = None
        self.age_diag = None
        self.brain_weight = None
        self.overall = None
        self.braak = None
        self.mmse = None
        self.ratio = ABeta42 / pTau
        Patient.all_patients.append(self)

    def __repr__(self):
        return f"{self.DonorID} | APOE Genotype: {self.apoe_genotype}
```

```
| ABeta40 {self.ABeta40} | ABeta42 {self.ABeta42} | tTau {self.tTau} |
pTau {self.pTau} | Death Age {self.death_age} | Cognitive Status
{self.cog_stat} | Age at Diagnosis {self.age_diag} | Braak
{self.braak} | Overall Change: {self.overall}"

    def get_id(self):
        return self.DonorID

    def get_ABeta40(self):
        return self.ABeta40

    def get_ABeta42(self):
        return self.ABeta42

    def get_ttau(self):
        return self.tTau

    def get_ptau(self):
        return self.pTau

    def get_cog(self):
        return self.cog_stat

    def get_death_age(self):
        return self.death_age

    def get_apoe(self):
        return self.apoe_genotype

    def get_age_diag(self):
        return self.age_diag

    def get_brain_weight(self):
        return self.brain_weight

    def get_overall(self):
        return self.overall

    def get_ratio(self):
        return self.ratio

    def get_braak(self):
        return self.braak

    def get_mmse(self):
        return self.mmse


    @classmethod
    def combine_data(cls, filename: str):
```

```
            with open(filename, encoding="utf8") as f:
                reader = csv.DictReader(f)
                rows_of_patients = list(reader)
                #for line in csv create object
                for row in range(len(rows_of_patients)):
                    if Patient.all_patients[row].DonorID ==
rows_of_patients[row]["Donor ID"]:
                        if rows_of_patients[row]["APOE Genotype"] !=
"":
                            Patient.all_patients[row].apoe_genotype =
rows_of_patients[row]["APOE Genotype"]

                        if rows_of_patients[row]["Age at Death"] !=
"":
                            Patient.all_patients[row].death_age =
int(rows_of_patients[row]["Age at Death"])

                        if rows_of_patients[row]["Cognitive Status"] !
= "":
                            Patient.all_patients[row].cog_stat =
rows_of_patients[row]["Cognitive Status"]

                        if rows_of_patients[row]["Age of Dementia
diagnosis"] != "":
                            Patient.all_patients[row].age_diag =
int(rows_of_patients[row]["Age of Dementia diagnosis"])

                        if rows_of_patients[row]["Fresh Brain Weight"]
!= "":
                            Patient.all_patients[row].brain_weight =
rows_of_patients[row]["Fresh Brain Weight"]

                        if rows_of_patients[row]["Overall AD
neuropathological Change"] != "":
                            Patient.all_patients[row].overall =
rows_of_patients[row]["Overall AD neuropathological Change"]

                        if rows_of_patients[row]["Braak"] != "":
                            Patient.all_patients[row].braak =
rows_of_patients[row]["Braak"]

                        if rows_of_patients[row]["Last MMSE Score"] !=
"":
                            Patient.all_patients[row].mmse =
rows_of_patients[row]["Last MMSE Score"]

                    else:
                        warnings.warn("IDs do not match.")

    @classmethod
```

```
    def instantiate_from_csv(cls, filename: str, other_file: str):
    #open csv and create list of all rows
        with open(filename, encoding="utf8") as f:
            reader = csv.DictReader(f)
            rows_of_patients = list(reader)
            #for line in csv create object
            for row in rows_of_patients:
                Patient(
                    DonorID = row['Donor ID'],
                    ABeta40 = float(row['ABeta40 pg/ug']),
                    ABeta42 = float(row['ABeta42 pg/ug']),
                    tTau = float(row['tTAU pg/ug']),
                    pTau = float(row['pTAU pg/ug'])
                )
            Patient.all_patients.sort(key = Patient.get_id)
            Patient.combine_data(other_file)

Patient.instantiate_from_csv('UpdatedLuminex.csv',
'UpdatedMetaData.csv')

# Creates a list of values related to it's respective column
cog = []
for patient in Patient.all_patients:
    cog.append(patient.get_cog())

age_death = []
for patient in Patient.all_patients:
    age_death.append(patient.get_death_age())

brain_weight = []
for patient in Patient.all_patients:
    if patient.get_brain_weight() == "Unavailable":
        brain_weight.append(None)
    else:
        brain_weight.append(int(patient.get_brain_weight()))

braak = []
for patient in Patient.all_patients:
    braak.append(patient.get_braak())

mmse = []
for patient in Patient.all_patients:
    mmse.append(patient.get_mmse())

# Creates a dictionary to match donor IDs to ratios
id_to_dict = {patient.get_id(): [patient.get_ABeta40(),
                                 patient.get_ABeta42(),
                                 patient.get_ptau(),
                                 patient.get_ttau(),
                                 patient.get_ratio(),
```

```
                                    patient.get_cog(),
                                    patient.get_death_age(),
                                    patient.get_brain_weight(),
                                    patient.get_overall(),
                                    patient.get_apoe(),
                                    patient.get_braak(),
                                    patient.get_mmse()] for patient in
Patient.all_patients}

# Fixes the data to date error in excel file
for patient in id_to_dict:
    if id_to_dict[patient][9] == "3-Mar":
        id_to_dict[patient][9] = "3/3"
    if id_to_dict[patient][9] == "4-Mar":
        id_to_dict[patient][9] = "3/4"
    if id_to_dict[patient][9] == "4-Apr":
        id_to_dict[patient][9] = "4/4"
    if id_to_dict[patient][9] == "3-Feb":
        id_to_dict[patient][9] = "2/3"
    if id_to_dict[patient][9] == "4-Feb":
        id_to_dict[patient][9] = "2/4"
```

```
Requirement already satisfied: scikit-posthocs in c:\users\hudso\
anaconda3\lib\site-packages (0.11.4)
Requirement already satisfied: numpy in c:\users\hudso\anaconda3\lib\
site-packages (from scikit-posthocs) (2.1.3)
Requirement already satisfied: scipy>=1.9.0 in c:\users\hudso\
anaconda3\lib\site-packages (from scikit-posthocs) (1.15.3)
Requirement already satisfied: statsmodels in c:\users\hudso\
anaconda3\lib\site-packages (from scikit-posthocs) (0.14.4)
Requirement already satisfied: pandas>=0.20.0 in c:\users\hudso\
anaconda3\lib\site-packages (from scikit-posthocs) (2.2.3)
Requirement already satisfied: seaborn in c:\users\hudso\anaconda3\
lib\site-packages (from scikit-posthocs) (0.13.2)
Requirement already satisfied: matplotlib in c:\users\hudso\anaconda3\
lib\site-packages (from scikit-posthocs) (3.10.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\
hudso\anaconda3\lib\site-packages (from pandas>=0.20.0->scikit-
posthocs) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\hudso\
anaconda3\lib\site-packages (from pandas>=0.20.0->scikit-posthocs)
(2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\hudso\
anaconda3\lib\site-packages (from pandas>=0.20.0->scikit-posthocs)
(2025.2)
Requirement already satisfied: six>=1.5 in c:\users\hudso\anaconda3\
lib\site-packages (from python-dateutil>=2.8.2->pandas>=0.20.0-
>scikit-posthocs) (1.17.0)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\hudso\
anaconda3\lib\site-packages (from matplotlib->scikit-posthocs) (1.3.1)
```

```
Requirement already satisfied: cycler>=0.10 in c:\users\hudso\
anaconda3\lib\site-packages (from matplotlib->scikit-posthocs)
(0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hudso\
anaconda3\lib\site-packages (from matplotlib->scikit-posthocs)
(4.55.3)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\hudso\
anaconda3\lib\site-packages (from matplotlib->scikit-posthocs) (1.4.8)
Requirement already satisfied: packaging>=20.0 in c:\users\hudso\
anaconda3\lib\site-packages (from matplotlib->scikit-posthocs) (24.2)
Requirement already satisfied: pillow>=8 in c:\users\hudso\anaconda3\
lib\site-packages (from matplotlib->scikit-posthocs) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\hudso\
anaconda3\lib\site-packages (from matplotlib->scikit-posthocs) (3.2.0)
Requirement already satisfied: patsy>=0.5.6 in c:\users\hudso\
anaconda3\lib\site-packages (from statsmodels->scikit-posthocs)
(1.0.1)
```

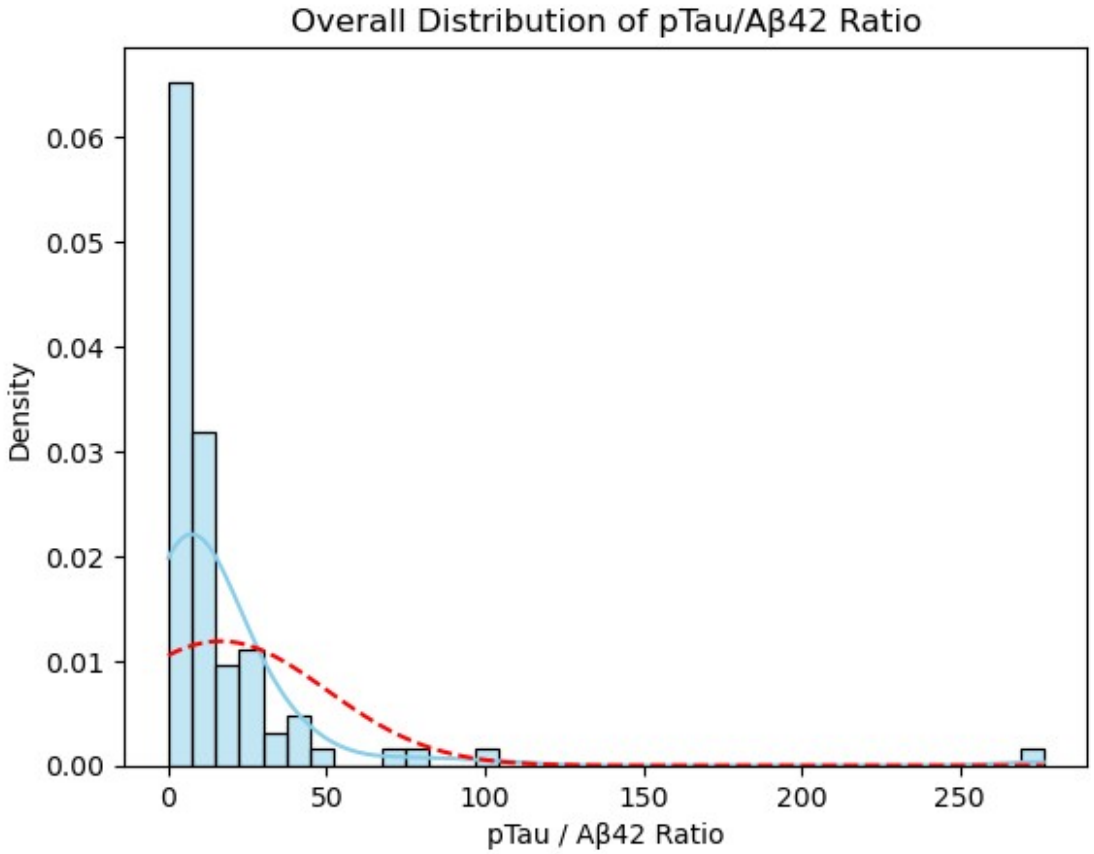
Data Analyis:

Checking Normality within the Data

- We are anaylzing how the ratio of pTau to amyloid beta 42 levels relates to the diagnosis and symptoms of Alzeihmer's Disease in order to gauge its effectiveness in doing so.
- Therefore, we will be comparing this ratio across various demographic data from our donors, requiring us to check it's normality for proper statistical tests.

```
import numpy as np
from scipy.stats import kruskal
import scikit_posthocs as sp
import pandas as pd
from scipy.stats import f_oneway
import seaborn as sns
from scipy.stats import norm, shapiro
import statistics

# Collect all ratios into one list
all_ratios = []
for patient in id_to_dict:
    ratio = id_to_dict[patient][4]
    if ratio is not None and ratio != "Unavailable":
        all_ratios.append(float(ratio))

all_ratios = np.array(all_ratios)

# Histogram with KDE and normal overlay
sns.histplot(all_ratios, kde=True, stat="density", color="skyblue",
edgecolor="k")
```

```
mean, stdev = np.mean(all_ratios), np.std(all_ratios)
x = np.linspace(min(all_ratios), max(all_ratios), 100)
plt.plot(x, norm.pdf(x, mean, stdev), "r--", label="Normal fit")
plt.xlabel("pTau / Aβ42 Ratio")
plt.title("Overall Distribution of pTau/Aβ42 Ratio")
plt.figure()
plt.show()

# Shapiro–Wilk test
stat, p = shapiro(all_ratios)
print(f"Shapiro–Wilk W={stat:.3f}, p={p:.4e}")
```




```
<Figure size 640x480 with 0 Axes>

Shapiro–Wilk W=0.432, p=2.0262e-16
```

Interpretation of pTau:AB42 Ratio Normaility

- The histogram displays the density of ratio values, indicating the mean and varaince between values. Projected onto the graph, the blue line represents the probability density function of the ratio values whereas the red, dashed line represents the normal distribution of the mean ratio values. Since these lines have little overlap, it is clear that our data is not normal and parametric tests can not be used to analyze the data.

- To confirm further, the Shapiro-Wilk Test mathmatically caluclates the normaility of the ratio values; with an indicated p-value of 2.4478e-16, much lower than the critical point of p = 0.05, it is clear that the data rejects the null hypothesis and is not normal.
- When analyzing our data, we are comparing the ratio of pTau to AB42 levels against demographics with multiple variables, requiring an ANOVA type test for comparision analysis. Since we concluded that the ratio data was not normal, a non-parametric ANOVA test will be used: Kruskal-Wallis Test.

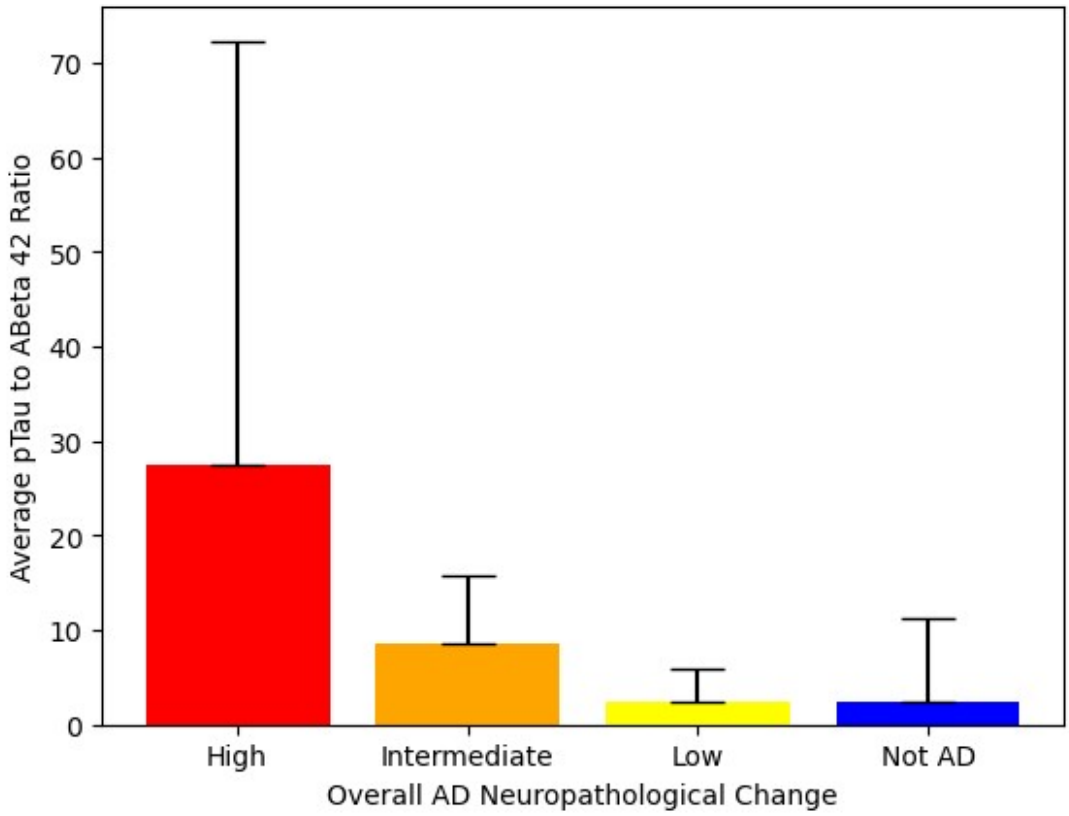
Comparison of pTau:AB42 Ratios to Overall AD Neuropathological Change (ADNC)

```
import numpy as np
from scipy.stats import kruskal
import scikit_posthocs as sp
import pandas as pd
from scipy.stats import f_oneway
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import norm, shapiro

# Creates list for values needed in comparison
high_to_ratio = []
intermediate_to_ratio = []
low_to_ratio = []
not_ad_to_ratio = []

# Filters the patients and sorts them depending on their Overall AD
neuropathological Change
for patient in id_to_dict:
    if id_to_dict[patient][8] == "High":
        high_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][8] == "Intermediate":
        intermediate_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][8] == "Low":
        low_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][8] == "Not AD":
        not_ad_to_ratio.append(id_to_dict[patient][4])

# Takes the mean of values from each category in order to reatain the
same length of data points
mean_high_to_ratio = statistics.mean(high_to_ratio)
mean_intermediate_to_ratio = statistics.mean(intermediate_to_ratio)
mean_low_to_ratio = statistics.mean(low_to_ratio)
mean_not_ad_to_ratio = statistics.mean(low_to_ratio)

# Takes the standard deviation from each category
sd_high_to_ratio = statistics.stdev(high_to_ratio)
sd_intermediate_to_ratio = statistics.stdev(intermediate_to_ratio)
sd_low_to_ratio = statistics.stdev(low_to_ratio)
```

```
sd_not_ad_to_ratio = statistics.stdev(not_ad_to_ratio)

# Defines the x and y columns used in the bar graph
overall_columns = ["High", "Intermediate", "Low", "Not AD"]
ratios = [mean_high_to_ratio, mean_intermediate_to_ratio,
mean_low_to_ratio, mean_not_ad_to_ratio]
stdevs = [sd_high_to_ratio, sd_intermediate_to_ratio, sd_low_to_ratio,
sd_not_ad_to_ratio]
yerr = [np.zeros(len(ratios)), stdevs]

# Perform the Kruskal-Wallis Test
h_stat, p_value = kruskal(high_to_ratio, intermediate_to_ratio,
low_to_ratio, not_ad_to_ratio)
print(f"Kruskal-Wallis H statistic: {h_stat:.3f}")
print(f"P-value: {p_value:.4e}")

alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: At least one group differs
significantly.")
else:
    print("Fail to reject the null hypothesis: No significant
difference between groups.")

# Dunn's post-hoc test
all_ratios = high_to_ratio + intermediate_to_ratio + low_to_ratio +
not_ad_to_ratio
group_labels = (['High'] * len(high_to_ratio) +
                ['Intermediate'] * len(intermediate_to_ratio) +
                ['Low'] * len(low_to_ratio) +
                ['Not AD'] * len(not_ad_to_ratio))

df = pd.DataFrame({'Ratio': all_ratios, 'Group': group_labels})

dunn_result = sp.posthoc_dunn(df, val_col='Ratio', group_col='Group',
p_adjust='holm')
print("\nDunn’s post-hoc test results:")
print(dunn_result)

## Plot the bar graph
plt.bar(overall_columns, ratios, yerr = yerr, capsize = 10, color =
["Red", "Orange", "Yellow", "Blue"])
plt.xlabel("Overall AD Neuropathological Change")
plt.ylabel("Average pTau to ABeta 42 Ratio")
plt.show()
```

```
Kruskal-Wallis H statistic: 34.093
P-value: 1.8933e-07
Reject the null hypothesis: At least one group differs significantly.
```

```
Dunn’s post-hoc test results:
                 High  Intermediate       Low    Not AD
High         1.000000      0.050459  0.000026  0.000056
Intermediate 0.050459      1.000000  0.052436  0.052436
Low          0.000026      0.052436  1.000000  0.804196
Not AD       0.000056      0.052436  0.804196  1.000000
```



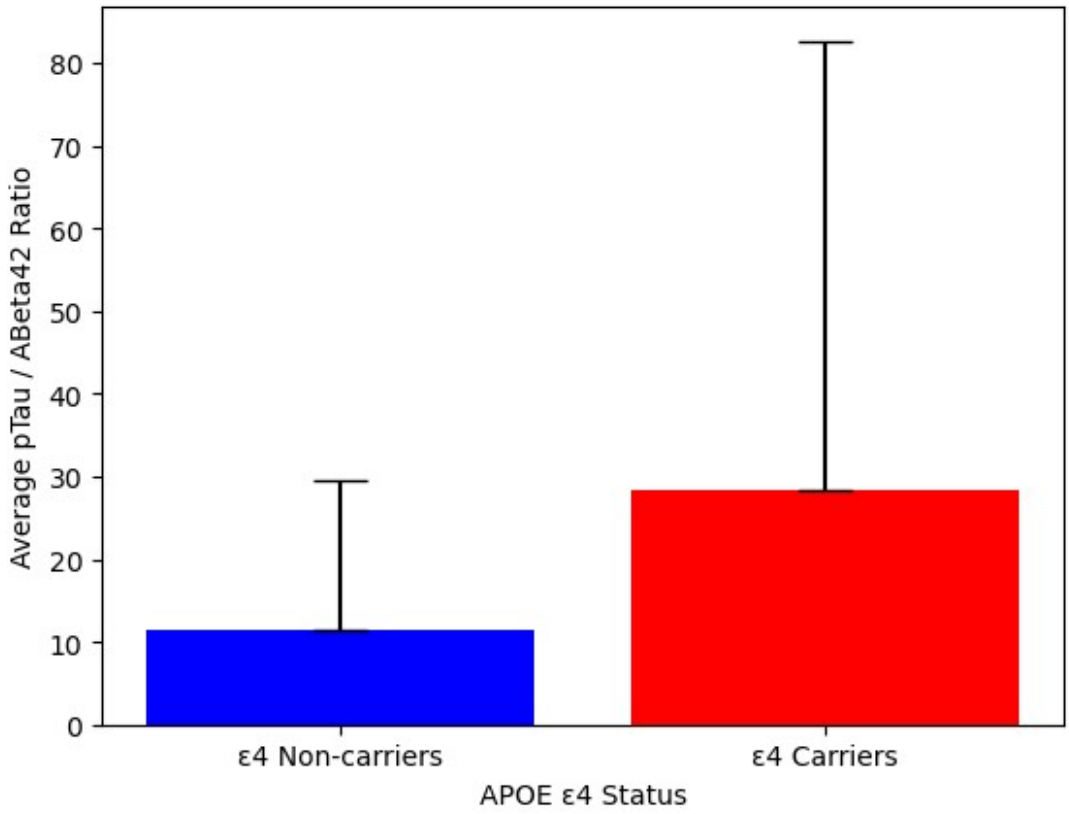
Interpreting of the Comparison of pTau:AB42 to Overall ADNC

- Kruskal–Wallis indicated significant differences in pTau/ABeta42 ratio across ADNC groups (H = 34.093, p = 1.8933e-07).
- Dunn’s post-hoc tests (Holm corrected) revealed that the High ADNC group differed significantly from both the Low (p < 0.001) and Not AD groups (p < 0.001). The comparison between High and Intermediate was marginal (p = 0.050). No other group differences were significant.
- The pTau:AB42 ratio levels within the High Overall ADNC group were also high themseleves, pointing toward a decrease in the amyloid beta-42 levels and an increase in the p-Tau levels as the patient's Overall ADNC increases.

Comparison of pTau:AB42 Ratios to APOE Geneotype-4 carriers and non-carriers

- The Mann-Whitney U Test is utilized as a non-parametric t-test to analayze the differences between carriers and non-carriers.

```
from scipy.stats import mannwhitneyu
import statistics
import matplotlib.pyplot as plt
import numpy as np

# Separate carriers vs non-carriers
carriers = []
noncarriers = []

for patient in id_to_dict:
    genotype = id_to_dict[patient][9]
    ratio = id_to_dict[patient][4]

    if "4" in genotype:
        carriers.append(ratio)
    else:
        noncarriers.append(ratio)

# Summary stats
mean_carriers = statistics.mean(carriers)
mean_noncarriers = statistics.mean(noncarriers)

sd_carriers = statistics.stdev(carriers)
sd_noncarriers = statistics.stdev(noncarriers)

# Plot bar graph
labels = ["ε4 Non-carriers", "ε4 Carriers"]
means = [mean_noncarriers, mean_carriers]
stdevs = [sd_noncarriers, sd_carriers]
yerr = [np.zeros(len(means)), stdevs]

plt.bar(labels, means, yerr=yerr, capsize=10, color=["Blue", "Red"])
plt.xlabel("APOE ε4 Status")
plt.ylabel("Average pTau / ABeta42 Ratio")
plt.show()

# Mann-Whitney U test
u_stat, p_value = mannwhitneyu(noncarriers, carriers,
alternative='two-sided')
print(f"Mann-Whitney U = {u_stat:.3f}, p = {p_value:.4e}")

alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: Carriers and non-carriers
differ significantly.")
else:
    print("Fail to reject the null hypothesis: No significant
difference between carriers and non-carriers.")
```




```
Mann-Whitney U = 448.000, p = 4.6932e-03
Reject the null hypothesis: Carriers and non-carriers differ
significantly.
```

Interpretation of the Comparison of pTau:AB42 to APOE Genotype-4 cariers and non-carriers

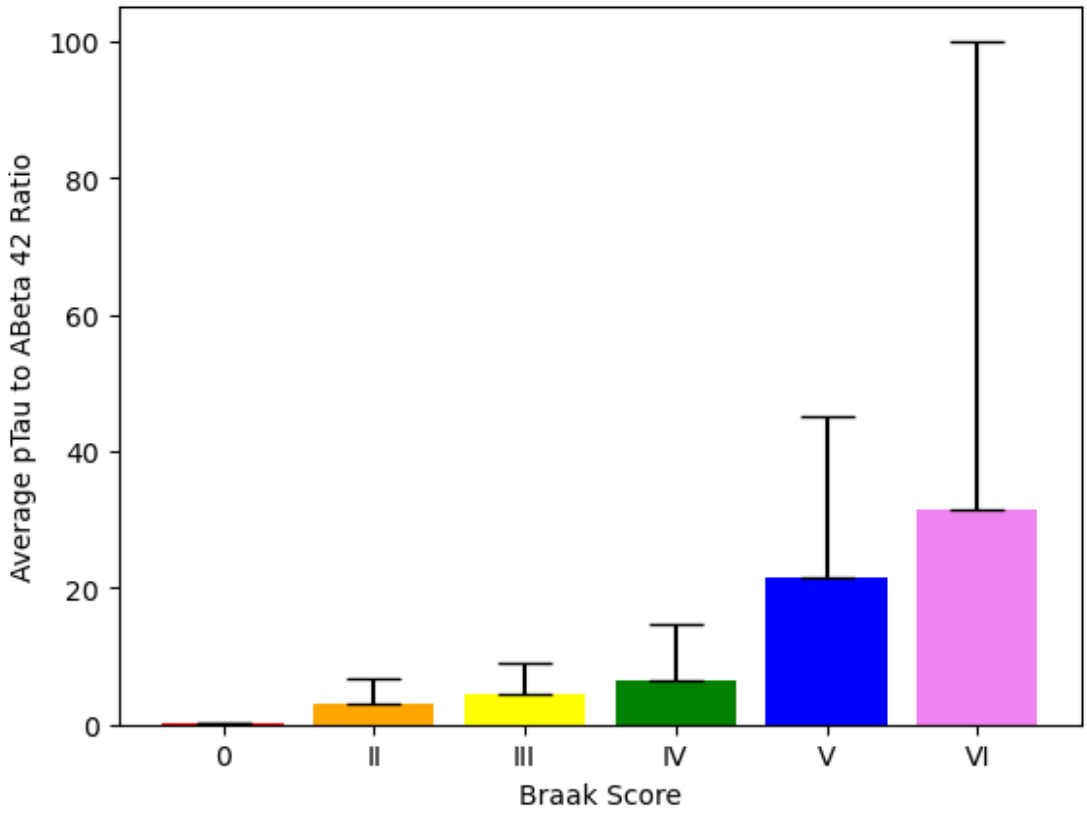
- Mann–Whitney U testing revealed a significant difference in pTau/Aβ42 ratios between APOE ε4 carriers and non-carriers (U = 448.0, p = 0.0047). Median ratios were higher in carriers than non-carriers, indicating that APOE ε4 status is associated with greater tau pathology relative to amyloid burden.

Comparison of pTau:AB42 Ratios to Braak Scores

```
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from scipy.stats import norm, shapiro
import numpy as np

braak_score = []
braak0_to_ratio = []
braakI_to_ratio = []
braakII_to_ratio = []
```

```
braakIII_to_ratio = []
braakIV_to_ratio = []
braakV_to_ratio = []
braakVI_to_ratio = []

for patient in id_to_dict:
    if id_to_dict[patient][10] == "Braak 0":
        braak0_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][10] == "Braak II":
        braakII_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][10] == "Braak III":
        braakIII_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][10] == "Braak IV":
        braakIV_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][10] == "Braak V":
        braakV_to_ratio.append(id_to_dict[patient][4])
    if id_to_dict[patient][10] == "Braak VI":
        braakVI_to_ratio.append(id_to_dict[patient][4])
    braak_score.append(id_to_dict[patient][10])


mean_braak0_to_ratio = statistics.mean(braak0_to_ratio)
mean_braakII_to_ratio = statistics.mean(braakII_to_ratio)
mean_braakIII_to_ratio = statistics.mean(braakIII_to_ratio)
mean_braakIV_to_ratio = statistics.mean(braakIV_to_ratio)
mean_braakV_to_ratio = statistics.mean(braakV_to_ratio)
mean_braakVI_to_ratio = statistics.mean(braakVI_to_ratio)

sd_braak0_to_ratio = statistics.stdev(braak0_to_ratio)
sd_braakII_to_ratio = statistics.stdev(braakII_to_ratio)
sd_braakIII_to_ratio = statistics.stdev(braakIII_to_ratio)
sd_braakIV_to_ratio = statistics.stdev(braakIV_to_ratio)
sd_braakV_to_ratio = statistics.stdev(braakV_to_ratio)
sd_braakVI_to_ratio = statistics.stdev(braakVI_to_ratio)

braak_columns = ["0", "II", "III", "IV", "V", "VI"]
ratios_braak = [mean_braak0_to_ratio, mean_braakII_to_ratio,
mean_braakIII_to_ratio, mean_braakIV_to_ratio, mean_braakV_to_ratio,
mean_braakVI_to_ratio]
stdevs = [sd_braak0_to_ratio, sd_braakII_to_ratio,
sd_braakIII_to_ratio, sd_braakIV_to_ratio, sd_braakV_to_ratio,
sd_braakVI_to_ratio]
yerr = [np.zeros(len(ratios_braak)), stdevs]

plt.bar(braak_columns, ratios_braak, yerr = yerr, capsize = 10, color
= ["Red", "Orange", "Yellow", "Green", "Blue", "Violet"])
plt.xlabel("Braak Score")
plt.ylabel("Average pTau to ABeta 42 Ratio")
plt.show()
```

```
# Kruskal-Wallis test
h_stat, p_value = kruskal(braak0_to_ratio, braakII_to_ratio,
braakIII_to_ratio,
                          braakIV_to_ratio, braakV_to_ratio,
braakVI_to_ratio)

print(f"Kruskal-Wallis H = {h_stat:.3f}")
print(f"P-value = {p_value:.4e}")

alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: At least one Braak group
differs significantly.")
else:
    print("Fail to reject the null hypothesis: No significant
difference between Braak groups.")

# Dunn's post-hoc test (if significant)
all_ratios = (braak0_to_ratio + braakII_to_ratio + braakIII_to_ratio +
              braakIV_to_ratio + braakV_to_ratio + braakVI_to_ratio)
group_labels = (['0'] * len(braak0_to_ratio) +
                ['II'] * len(braakII_to_ratio) +
                ['III'] * len(braakIII_to_ratio) +
                ['IV'] * len(braakIV_to_ratio) +
                ['V'] * len(braakV_to_ratio) +
                ['VI'] * len(braakVI_to_ratio))

df = pd.DataFrame({'Ratio': all_ratios, 'Group': group_labels})

dunn_result = sp.posthoc_dunn(df, val_col='Ratio', group_col='Group',
p_adjust='holm')
print("\nDunn's post-hoc test results:")
print(dunn_result)
```




```
Kruskal-Wallis H = 22.414
P-value = 4.3673e-04
Reject the null hypothesis: At least one Braak group differs
significantly.

Dunn's post-hoc test results:
            0        II       III        IV         V        VI
0    1.000000  1.000000  1.000000  1.000000  0.123709  0.161833
II   1.000000  1.000000  1.000000  1.000000  0.219467  0.333807
III  1.000000  1.000000  1.000000  1.000000  0.219467  0.333807
IV   1.000000  1.000000  1.000000  1.000000  0.014340  0.139748
V    0.123709  0.219467  0.219467  0.014340  1.000000  1.000000
VI   0.161833  0.333807  0.333807  0.139748  1.000000  1.000000
```

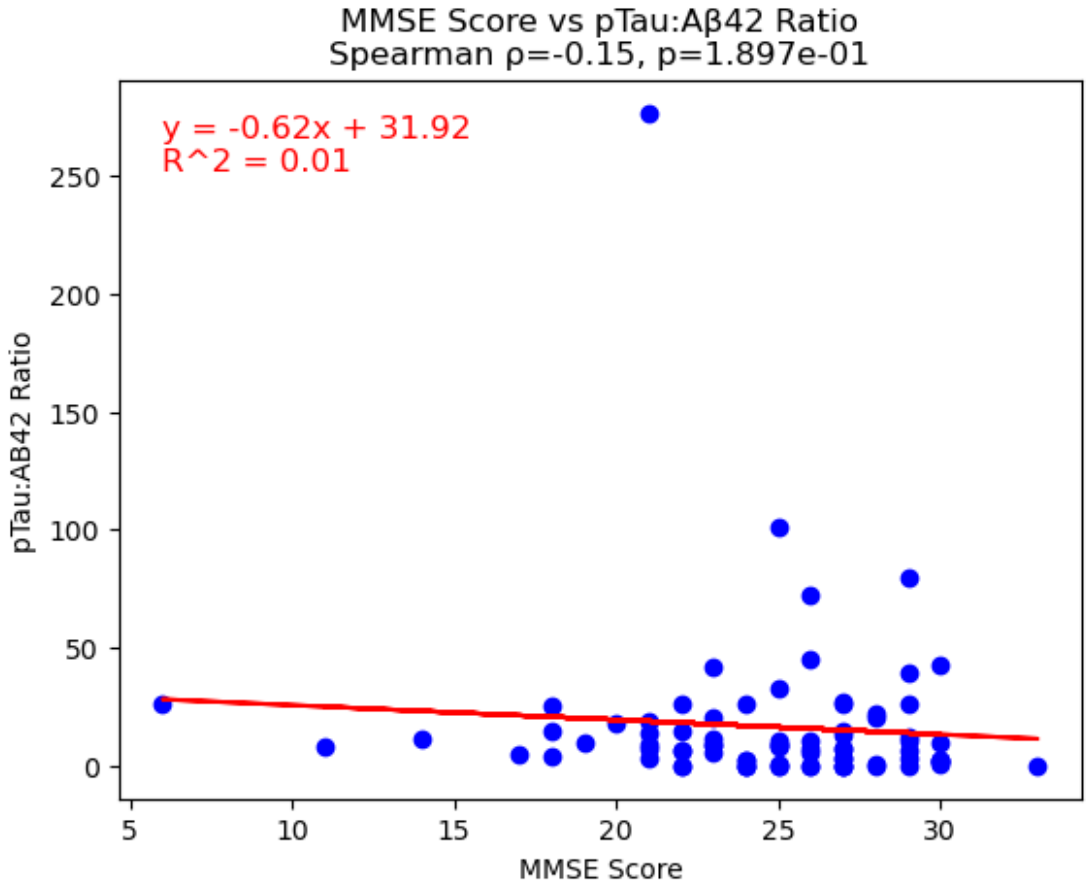
Interpretation of the Comparison of pTau:AB42 ratios to Braak Score

- Kruskal–Wallis testing revealed a significant effect of Braak stage on pTau/Aβ42 ratios ($H = 22.41$, $p < 0.001$).
- Dunn's post-hoc tests with Holm correction showed that Braak stage IV differed significantly from Braak stage V ($p = 0.014$), while no other pairwise comparisons reached significance. These findings suggest that progression from Braak IV to Braak V may be a critical point in biomarker changes.

Comparison of pTau:AB42 Ratios to MMSE Scores

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import spearmanr, linregress
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

# Creates a list for values needed in scatterplot
ratios = []
mmse_scores = []
for donor, values in id_to_dict.items():
    ratio = values[4]
    mmse  = values[11]

    if ratio is not None and mmse is not None:
        ratios.append(float(ratio))
        mmse_scores.append(float(mmse))

# Linear Regression and Scatterplot
df = pd.DataFrame({
    'MMSE Score': mmse_scores,
    'pTau:AB42': ratios})

df.to_csv('patient_data.csv', index = False)
print("CSV file 'patient_data.csv' has been created")

df = pd.read_csv("patient_data.csv")

x = df["MMSE Score"].values.reshape(-1,1)
y = df["pTau:AB42"].values

model = LinearRegression()
model.fit(x, y)

slope = model.coef_[0]
intercept = model.intercept_
r2 = model.score(x,y)

plt.scatter(x, y, color = 'blue')
plt.plot(x, model.predict(x), color = "red")

equation = f"y = {slope:.2f}x + {intercept:.2f}\nR^2 = {r2:.2f}"
plt.text(x.min(), y.max(), equation, color = 'red', fontsize = 12,
verticalalignment = 'top')
plt.xlabel("MMSE Score")
plt.ylabel("pTau:AB42 Ratio")

# Spearman correlation
rho, pval = spearmanr(mmse_scores, ratios)
```

```
plt.title(f"MMSE Score vs pTau:Aβ42 Ratio\nSpearman ρ={rho:.2f},
p={pval:.3e}")
plt.show()

CSV file 'patient_data.csv' has been created
```



Interpretation of the Comparison of pTau:AB42 Ratios to MMSE Scores

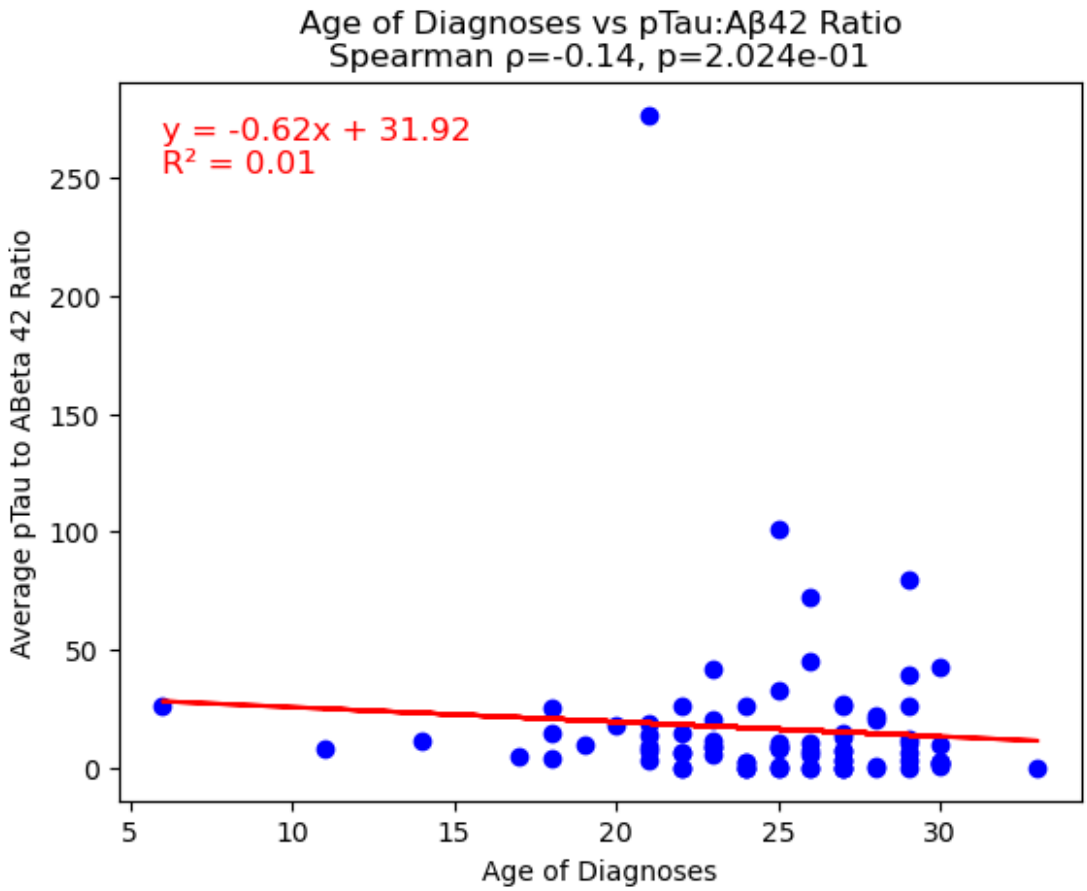
- Though the line of best fit has a negative slope, its significance determined by its R2 avlue and p-value indicate that there is no significant correlation betweent he two groups: R2 = 0.01 (far from a value of 1) and p ~ 0.19 (far above a value of 0.05).

Comparison of pTau:AB42 Ratios to Age of Diagnoses

```
#Create lists of ages of diagnoses for patients and ratios of pTau to
AB42 ratios for patients
age_diag = []
ratios = []

for patient in id_to_dict:
    if id_to_dict[patient][11] != None:
        age_diag.append(id_to_dict[patient][11])
```

```
        ratios.append(id_to_dict[patient][4])

# Export data as a CSV File
import pandas as pd
print(age_diag)
print(ratios)

# Create a DataFrame
df = pd.DataFrame({
'Age of AD Diagnoses': age_diag,
'Average pTau to ABeta 42 Ratio': ratios
})

# Write to CSV
df.to_csv('patient_data2.csv', index=False)
print("CSV file 'patient_data2.csv' has been created.")

# Run Linear Regression on data in patient_data file
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
df = pd.read_csv("patient_data2.csv")
x = df["Age of AD Diagnoses"].values.reshape(-1, 1)
y = df["Average pTau to ABeta 42 Ratio"].values
model = LinearRegression()
model.fit(x, y)
slope = model.coef_[0]
intercept = model.intercept_
r2 = model.score(x, y)

plt.scatter(x, y, color='blue')
plt.plot(x, model.predict(x), color="red")

# Annotate equation
equation = f"y = {slope:.2f}x + {intercept:.2f}\nR² = {r2:.2f}"
plt.text(x.min(), y.max(), equation, color="red", fontsize=12,
verticalalignment='top')

# Annotate scatterplot with labels and title
plt.xlabel("Age of Diagnoses")
plt.ylabel("Average pTau to ABeta 42 Ratio")

# Spearman correlation
rho, pval = spearmanr(age_diag, ratios)
plt.title(f"Age of Diagnoses vs pTau:Aβ42 Ratio\nSpearman ρ={rho:.2f},
p={pval:.3e}")
plt.show()

['25', '28', '33', '25', '29', '29', '21', '29', '25', '27', '25',
'21', '30', '14', '28', '22', '23', '22', '23', '25', '26', '27',
'25', '30', '24', '29', '18', '23', '26', '27', '25', '26', '24', '6',
```

```
'18', '25', '27', '27', '30', '24', '26', '21', '11', '27', '25',
'30', '23', '28', '24', '22', '19', '23', '28', '23', '24', '30',
'24', '29', '30', '22', '29', '26', '27', '27', '30', '27', '18',
'27', '21', '26', '27', '20', '22', '29', '21', '29', '21', '17',
'22', '26']
[0.5110741968137156, 1.0023068049018278, 0.056257545365441745,
10.828315819363615, 12.172085650363682, 39.63088897702834,
7.0882860671112295, 10.496526168353485, 7.977634700809873,
15.108870965872951, 9.321955719971537, 276.40342181147673,
9.99014598631279, 11.804891302224945, 20.126788949409192,
25.94895922728319, 8.901470126872022, 6.4724919091419,
9.040897098009339, 8.882060333841464, 5.431414739076327,
3.5489153363977666, 33.22937341744784, 1.9123376619884125,
0.13011734032138966, 80.043585546875, 14.80580511425044,
20.199423482193357, 7.07189542409935, 7.017452009542461,
100.81233598222464, 10.541780448039553, 0.043559543857036476,
26.363284740931196, 4.396839041417106, 0.13335642540815468,
0.13104152487864124, 0.19291338586981405, 2.0931092439101135,
26.178678182077626, 44.8343842349808, 3.075558206762797,
8.12359072509184, 12.818181820877726, 0.044993103309569085,
0.9543889845211926, 41.92898643777547, 22.06188282114862,
2.016610596595348, 0.12840967577825993, 9.722616629859642,
5.505801685573976, 0.008884652214914381, 11.464579379344528,
2.018282989392277, 2.476368581709313, 0.06816760467977367,
3.2270457164233575, 42.38361266101955, 14.703185705493643,
0.06667809074290254, 71.99122808823529, 3.5581583196328945,
25.958058838014683, 2.158565202866242, 7.118424155831036,
25.2166748972111, 27.036500691760846, 13.58746933269709,
0.19259092412951268, 0.04118062921308692, 18.38410764376056,
0.127392017543609, 26.4982600434435, 19.116708649015884,
6.7872848955631415, 9.185398104161486, 4.689418233647062,
6.133871509494374, 0.013717986575179452]
CSV file 'patient_data2.csv' has been created.
```



Interpretation of the Comparison of pTau:AB42 Ratios to Ages of Diagnoses:

- Though the line of best fit has a negative slope, its significance determined by its R2 avlue and p-value indicate that there is no significant correlation betweent he two groups: R2 = 0.01 (far from a value of 1) and p ~ 0.202 (far above a value of 0.05).

Verify and validate your analysis:

- Throughout our analysis, the pTau:AB42 ratios of donors showed significanct difference between the maximum and minimum values of ranked demographic data from the Luminex data set; however, there failed to be any significant difference between the maximum and minimum values compared to the intermediary counterpart. This is shown statistacally via Dunn's post-hoc test on the comparison between pTau:AB42 ratios and Overall ADNC; Mattsson-Carlgren Et al.'s research on CSF Biomarkers in Autopsy-Confrimed AD and Frontotemporal Lobar Degernration explores this comparison further, showing that confirmed AD donors within the "High" to "Intermediate" range of Overall ADNC had, on average, higher ratios compared to donors within the "Low" to "Not AD" range. Though there is little difference between the Intermediary ADNC and the High/Low/Not AD ADNC, our verified analysis points to a correlation between the pTau:AB42 ratios and Overall ADNC.
  - https://www.neurology.org/doi/10.1212/WNL.0000000000200040

- Researched by Troutwine Et al., the effect of the APOE-4 Genotype offers a significant increased risk for AD; this is in part due to APOE-4's lack of effectiveness when clearing amyloid beta-42 from the brain and its excessive tau hyperphosphorylation, spreading tau-toxicity across the brain. Using this as a foundation, the pTau:AB42 ratios were compared to non-carriers and carriers of the APOE-4 genotype, showing significant difference between the groups and a rise in pTau:AB42 ratio levels in carriers of APOE-4 genotype carriers. Benson Et al. explores this increase in pTau:AB42 ratios compared to APOE-4 genotype carriers and concludes that APOE-4 genotype carriers have a decrease in amyloid beta-42, verifying our analysis and suggesting that APOE-4 genotype has the ability to modulate biomarkers within the human body.
  - https://www.sciencedirect.com/science/article/pii/S2211383521003944
  - https://www.researchgate.net/publication358017720_Don%27t_forget_about_tau_the_effects_of_ApoE4_genotype_on_Alzheimer%27s_disease_cerebrospinal_fluid_biomarkers_in_subjects_with_mild_cognitive_impairment-data_from_the_Dementia_Competence_Network
- Our analysis falls short, besides the intermediaries of ranked demographic data, when determing the correlation between pTau:AB42 ratios and Braak and MMSE Scores. Throughout both scatterplot graphs, the R^2 values and p-values show no significance in correlation. When considering the dynamic between pTau:AB42 ratios and Braak scores, the lack of correlation can be explained based on when biomarker levels were taken. On the other hand, the lack of correlation between pTau:AB42 ratios and MMSE scores reflects the research of Se-Hwee Oh Et al.: biomarkers explain only 20-40% of variance in cognitive impairment, highlighting the numerous non-pathological factors that affect not only the cognitive status of the donors but also their biomarkes.
  - https://www.nature.com/articles/s41591-025-03565-2
- In verifying our analysis of data with statistical tests and published papers, our hypothesis is proven partially true. Though pTau:AB42 ratio levels can refelct AD symptoms, it does not provide causation/correlation to all demographics. Moreover, pTau:AB42 ratio comparisons reflect pathology based variances (APOE genotypes and Overall ADNC) over clinical based variances in which demogrpahics are assigned to the donor (Braak and MMSE Scores).

Conclusions and Ethical Implications:

- Throughout this analysis, we have explored numerous comparisons between pTau:AB42 ratios and demographic data provided from a cohort study on 84 donors. In this analysis, statistical test results and published research papers point toward pTau:AB42 ratios as a pathology-anchored biomarker, as seen in the Overall ADNC and APOE-4 genotype comparisons, leaving pTau:AB42 ratios as a poor indicator of AD sympotms in clincal-based demogrpahics. Through our partially supported hypothesis, ethical concerns are raised regarding the utilization of pTau:AB42 ratio levels as indicators of AD via pathology-based comparisons. With one of our comparisons relying on APOE-4 genotype carriers having an increased risk in developing AD, our analysis overinterprets and stigmatizes carriers into a AD diagnoses, raising ethical implications regarding research techniques and analysis practices. Another implication revolves around the clinical use of biomarkers as the sole marker in AD diagnoses, overgeneralizing the amount of factors affecting AD diagnoses and stigmatizing patients with high biomarker levels. Without addressing these ethical concerns, the continuity of care and accurated

diagnoses for AD patients is only limited. In conclusion, our analysis of the effectiveness of pTau:AB42 ratio levels as an indicator for AD was proven to only show significance with pathological demographics, with wide variance and uncertainty regarding clincal demographics.

Limitations and Future Work:

- Throughout our analysis, there was broad variance of our pTau:AB42 ratio levels, reflected in each of our comparitive graphs and statistical tests, limiting the true value of significance bewtween each category tested. This variance could have been combated through filtering out the outliers within our data set; however, due to the Meta and Lumiex data set sample size, this was not feasible as every value counted. In the future, a larger sample size would allow for less variance and the ability in filtering out any outliers. This larger sample size would also increase the amount of donors with more of their demographics "filled in"; moreover, numerous donors have null values within their respective columns, indicating tha lack of data aquired and decreasing the overall viablity of that donor when comparing to pTau:AB42 ratio levels. Looking toward the future, the effectiveness of pTau:AB42 ratio levels can be further explored through the comparison of other biomarker ratio levels and against a control group where AD is not present. Further tests would only gauge the effectiveness when relating to pathological demographics, as there is too much variability within a donor's cognitive status to gauge effectiveness with clinical demographics.