Module 1:

Team Members:

Logan Heselton and Luke Shee

Project Title:

Alzheimer's Disease: Identifying Relationships between Beta-amyloid-40 vs. Beta-amyloid-42 and AD Neuropathological Change

Project Goal:

This project seeks to examine whether overall AD neoropathological change (OADC) has a stronger correlation with betaamyloid-40 or betaamyloid-42 concentration in American Alzheimer's patients. To guide this, it aims to answer the following questions:

1. Are betaamyloid-40 and betaamyloid-42 concentrations correlated with each other?
2. Does overall AD neuropathological change have a higher correlation with betaamyloid-40 or betaamyloid-42 concentration in American Alzheimer's patients?

Disease Background:

Prevalence & incidence

- Prevalence: ~6.9 million people currently living with Alzheimer's disease in the U.S. (varies by source, but sources suggest around 6.8-7.2 million people)
  - Source: https://pubmed.ncbi.nlm.nih.gov/38689398/
- Incidence: ~450,000-500,000 thousand new cases every year in the U.S.
  - Source: https://www.alzsd.org/resources/facts-stats/

Economic burden

- Timeline is from 2016 to 2060 (future estimations)
  - 2016
    - The estimated per-patient cost of formal care was $28,078 (confidence interval range$:25,893–$30,433)
    - The estimated per-patient cost of informal care (replacement cost and forgone wages) was $36,667 (confidence interval$:34,025–$39,473) and 15,792 (confidence interval: $12,980$–18,713), respectively.

        - Source: https://www.nature.com/articles/s41514-024-00136-6
    - 2020
        - Aggregate formal + informal care costs using replacement cost and forgone wage methods were $196 billion (confidence interval: 179–213 billion), 450 billion (confidence interval: 424–478 billion), and $305 billion (confidence interval: 278–$333 billion), respectively
        - Source https://www.nature.com/articles/s41514-024-00136-6
    - 2025 (current)
        - Medical and long-term care for patients with dementia will cost the United States $232 billion this year, including 52 billion paid out of pocket by patients and their families.
            - More than two-thirds of the total cost of care is paid for by Medicare ($106 billion) and Medicaid (58 billion).
        - Source: https://schaeffer.usc.edu/research/dementia-alzheimers-cost-model-2025/
    - 2060
        - Expectation is that the above numbers in 2020 were going to increase to $1.4 trillion (confidence interval: 837 billion–$2.2 trillion), 3.3 trillion (confidence interval: 1.9–5.1 trillion), and $2.2 trillion (confidence interval: 1.3–$3.5 trillion), respectively.
        - Source: https://www.nature.com/articles/s41514-024-00136-6

Risk factors (genetic, lifestyle)

- Genetic:
    - Apolipoprotein E (APOE)
        - APOE e2: Least common form of the gene; reduces the risk for Alzheimers
        - APOE e4: More common form of the gene; increases the risk for Alzheimers; linked to more severe forms of Alzheimers
        - APOE e3: Most common form of the gene; doesn't seem to affect the risk for Alzheimers
    - A child whose biological parent carries a genetic variant for one of these three genes has a 50/50 chance of inheriting that altered version of the gene. If the variant is inherited, the child has a very strong probability of developing Alzheimer's before age 65 and sometimes much earlier.
    - Source: https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-causes-alzheimers-disease
- Lifestyle:
    - Age
        - Most individuals with the disease are 65 and older. After age 65, the risk of Alzheimer's doubles every five years. After the age of 85, the risk increases to nearly one-third.
    - Head injuries

        - There is a link between head injury and future risk of dementia.
    - Other factors:
        - Unmanaged chronic health issues, such as high blood pressure or hearing loss
        - Physical inactivity
        - Unhealthy diet
        - Alcohol misuse
        - Smoking
        - Not getting enough sleep or not sleeping well
        - Social isolation
        - Lack of mental stimulation
    - Source: https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-causes-alzheimers-disease

Societal determinants

- Education
    - Studies show an association between lower level of education and poorer brain health
        - Research suggests that the proportion of adults older than age 45 and older with worsening memory loss (or dementia)/frequent cognitive confusion have been linked to individuals who didn't receive a high school diploma and is unassociated with college graduates.
- Access to Health Care
    - Dementia risk can be reduced with continued access to health care; it can:
        - Help people prevent chronic diseases by identifying behaviors and conditions that can lead to serious health problems.
        - Aid in early diagnosis of many health conditions, such as diabetes, heart disease, and dementia.
        - Help people manage chronic conditions to avoid complications and hospitalizations.
        - Help diagnose dementia earlier to allow for better care coordination and support.
- Environment
    - A person's individual environment can affect their quality of life: their eating style, physical health, and housing.
    - Social isolation could also increase a person's risk of premature death from all causes; people experiencing social isolation or loneliness are at a higher risk for dementia
- Economic Stability
    - The increased risk to unemployment are linked to loss of health care, depression, and unhealthy coping mechanisms
    - People who are employed and who are in better economic status are better positioned to use preventative services and consistently practice healthier practices.

- Source: https://www.cdc.gov/alzheimers-dementia/php/sdoh/index.html and https://brainhealth.dc.gov/page/social-determinants-health-and-dementia

Symptoms

- Early-stage symptoms:
  - Memory loss/Thinking and Reasoning
    - Memory loss becomes worse over the course of the development of the disease.
    - Contributes to: repeat statements and questions over and over, forget conversations, appointments or events, misplace items, often putting them in places that don't make sense, get lost in places they used to know well, forget the names of family members and everyday objects, have trouble finding the right words, expressing thoughts or having conversations.
    - Concentrating is difficult for people with Alzheimer's, especially for long periods of time. Doing more than one task at once is especially hard.
- Mid-stage symptoms:
  - Changes in Personality and Behavior
    - Brain changes can affect mood and behaviors: depression, loss of interest in activities, social withdrawal, mood swings, not trusting others, anger or aggression, changes in sleeping habits, wandering, loss of inhibitions, and delusions, such as believing something has been stolen when it hasn't.
- Late-stage symptoms:
  - In the late stages of the disease, the symptoms become increasingly severe and can be distressing for the person with the condition, as well as their carers, friends and family.
    - Severe mood changes: be aggressive, such as hitting or shouting, become upset or restless, and call out or repeat the same question again and again
    - Other signs include: difficulty eating and swallowing (dysphagia), unintentional passing of urine (urinary incontinence) or stools (bowel incontinence), difficulty changing position or moving around without assistance, and loss of speech.
    - In the severe stages of Alzheimer's disease, people may need full-time care and assistance with eating, moving and personal care.
- Source: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447 and https://www.nhs.uk/conditions/alzheimers-disease/symptoms/

Diagnosis

- Doctors and healthcare experts mainly diagnose Alzheimer’s disease in patients by interrogating them on their dementia-like symptoms and by running a variety of examinations on their cognitive ability and body chemistry.

    - Likely a traditional first step to diagnosing Alzheimer's is noticing its symptoms, including mood changes, poor judgement, trouble doing daily tasks, and others mentioned above. In addition, a doctor, neurologist or geriatrician may also ask for more details about the patients' symptoms and medical history to further evidence the Alzheimer's diagnosis and learn its potential cause. They may also interview the patient's family members and close colleagues for more information about the patient's condition.
    - To gain confidence in the diagnosis, the health care team may run cognitive and physical examinations on their patients. Specialists can give patients mental status and/or neuropsychological tests to more objectively judge their memory and cognitive ability. Moreover, doctors can execute pre-established tests for other health issues like depression, Parkinson's disease, or vitamin B-12 deficiency to double check that the cognitive symptoms don't come from a different condition. Lastly, Alzheimer's patients may be asked to go through brain imaging to check for brain degeneration that Alzheimer's may cause; while brain scans can't confirm Alzheimer's diagnosis, they can still rule out more alternative diseases. More reliable diagnosis measures for Alzheimer's are still being researched, including some that look into amyloid or tau-protein levels in blood and biomarkers like genes.
- Source: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers/art-20048075

Standard of care treatments (& reimbursement)

- Treatments include drugs that regulate cognitive and behavioral symptoms (e.g. Galantamine, benzgalantamine, rivastigmine, and donepezil are cholinesterase). Other FDA-approved drugs like lecanemab and donanemab reduce beta-amyloid plaques in early Alzheimer's stage to slow the cognitive decline (there's also some insurance to cover the drug costs).
- In moderate to severe cases, memamtine and other medications can be prescribed to regulate glutamate in the brain, reducing Alzheimer's symptoms to allow patients to be able to do more ordinary tasks, like going to the bathroom independently.
    - Dosage of such medications vary based on how well patients tolerate them, starting with low doses and gradually increasing if tolerated. They also have side effects, including nausea, vomiting, confusion, headaches, dizziness, loss of appetite, etc.
    - Source: https://www.nia.nih.gov/health/alzheimers-treatment/how-alzheimers-disease-treated
- Most of Alzheimer's care can be paid for with Medicare primarily and other private health care insurance for those 65+ years old, at least for initial costs. Medicaid and long-term plans help cover Alzheimer's expenses in the long term.
    - Source: Google AI and https://www.alz.org/help-support/caregiving/financial-legal-planning/paying-for-care

Disease progression & prognosis

- General 5 stages of Alzheimer's (generally lasts 3-20+ years)
    1. Preclinical Alzheimer's:
        - Stage long before noticeable symptoms, usually only identified in research settings, that can last for many years. Tau proteins (total Tau and phosphorylated Tau) aggregate into tangles within neurons, hindering their functioning, and Beta-amyloid plaques grow between synapses to disturb neuron communication. Patients start taking tests to see signs of Alzheimer's risk, which include blood sample biomarkers, brain imaging, and genetic tests.
    2. Mild cognitive impairment due to Alzheimer's:
        - Patients start having mild impairments in memory and thinking ability, such as memory lapses during conversation/events and trouble judging or planning tasks. Doesn't have much effect on work or relationships.
    3. Mild dementia due to Alzheimer's:
        - This is when the patient shows difficulties doing daily tasks due to signs of dementia, mainly bad memory and/or cognitive capacity (possibly some personality changes to be more withdrawn or irritable). Alzheimer's is likely to be diagnosed at this stage.
    4. Moderate dementia due to Alzheimer's:
        - The patient's dementia worsens, amplifying the prior symptoms and calling for the need of a care person to help the patient do some daily activities. They also keep growing more irritable and unnecessarily suspicious about close ones.
    5. Severe dementia due to Alzheimer's:
        - Patient starts losing the ability to communicate and execute regular physical activities like walking, sitting, and eventually swallowing and bowel functioning.
- Source: https://www.alz.org/alzheimers-dementia/stages?utm_source=google-grant&utm_medium=paidsearch&utm_campaign=google_grant&gad_source=1&gad_camp 96Sk0O-LVITT-FHXOO6JRKsXFyOWi1aikNi0iCyiQFy2UsGQRoCQXgQAvD_BwE

Continuum of care providers

1. Early Stage

- For the Alzheimer's patients in America, continuum of care starts with the diagnosis. The supportive care team usually includes a primary care physician or specialist, to provide medical & symptom knowledge and management, as well as a social worker from a nonprofit organization or community service, to assist in Alzheimer's guidance, resources, and coping for the family.
    - Specialists may include neurologists (oversee disorders in brain and nervous system), geriatricians (cares for older adults), geriatric psychiatrists (gives mental and emotional health care for older adults), neuropsychologists (specializes in standardized assessment of cognition and behavior, verbatim from gov source),

speech and physical therapists (helps improve patients' communication and activity abilities)

2. Mid Stage

- The Alzheimer's symptoms start progressing, so patients start needing outside help for their everyday activities. Although family members may play the caregivers for the patients, other hired care services like visiting nurses and home health aides can assist in tasks like toileting, dressing, bathing, shopping, cooking, and light housekeeping for the Alzheimer's patients. Caring family members may also hire short-term respite care to temporarily support the patients while the primary caregivers take a break (alzoc).Caregivers can also consider sending patients to Adult Day Care, which offers supervised physical and social activities as medical treatment and therapeutic recreation.

3. Late Stage

- By this phase, the symptoms of the Alzheimer's patient may make them no longer able to stay at home safely due to difficulty swallowing or other dementia-induced accidents. The supportive care team and/or close family members may transition the patient to hospice (end-of-life) care, nursing home, or other long-term residential facility to provide them 24/7 medical care and supervision.

- Source: Google AI, https://www.alzheimers.gov/professionals/health-care-providers, and https://www.hospicechesapeake.org/2022/06/continuum-of-care-offers-dementia-patients-better-quality-of-life/chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.alzoc.org/wp-content/uploads/2015/11/RESPITE-Continuum-of-Care-AlzOC-1.2016-exp-6.2016.pdf

Biological mechanisms (anatomy, organ physiology, cell & molecular physiology)

- Alzheimer's is a disease targeting the brain, namely its neuron cells that are key to its functionality. Neurons work through communication: a neuron receives cellular signals, in the form of neurotransmitter chemicals sent by other neurons binding to the receptors within the branch-like dendrites of the said neuron, that causes it to send an electric charge down through its cable-like axon neck and send out more neurotransmitters to signal other neurons in the in-between space, called synapses. This communication among millions of neurons, with the aid of consistent glucose sustenance, astrocyte and microglia support, and brain usage, eventually results in humans' everyday cognitive abilities.
- Alzheimer's disease has multiple observed brain changes that may lead up to and cause the condition, such as…
    - Beta-amyloid plaques: beta-amyloid is a protein that clumps in the brain around neurons, namely in-between neuron synapses that interferes with their cellular

communication. Too much of these plaques, especially with beta-amyloid 42 form, can greatly hinder neuron functionality and cause cognitive decline.
    - Tau tangles: tau is another protein that usually stabilizes microtubules in healthy neurons. In Alzheimer's patients though, tau detaches from the microtubules and groups together to form neurofibrillary tangles inside neurons, perhaps spurred by overaccumulation of beta-amyloid plaques. These tangles then harm neurons by blocking their transport and communication systems from within.
    - Connection Loss and Dysfunction: As neurons stop functioning properly, neuron connections and networks start falling apart, causing brain regions to shrink and cognitive ability to decline.
    - Chronic Inflammation: initially-helpful glial cells that begin malfunctioning may induce inflammation in the brain due to lack of debris disposal and management.
    - Vascular Issues: since the brain relies heavily on blood and the glucose it provides to keep neurons healthy and active, cardiovascular issues hindering that blood glucose supply may lead to brain inflammation and later neural connection breakdowns.
- Source: https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease

Clinical Trials/next-gen therapies

- Clinical trials for Alzheimer's disease have been going on to engineer future drugs and therapies that reduce patients' symptoms and grant them more time to function independently without the need of a caregiver. Most of those developing drugs and therapies aim to fight against the main Alzheimer's protein clumps, betaamyloid plaques and tau tangles, but other Alzheimer's treatments like gene therapy are also being looked into.

    - For example, lecanemab and donanemab drugs, recently approved by the FDA (Food and Drug Administration) in the past few years, act as antibodies that neutralize betaamyloid proteins to prevent them from forming plaques. These drugs are estimated to give patients with mild symptoms several more months of independent living, specifically 8 additional months from donanemab or 10 additional months from lecanemab.
        - Source: https://medicine.washu.edu/news/next-gen-alzheimers-drugs-extend-independent-living-by-months/
- University medical institutes and other online Alzheimer's research organizations like Alzheimers.gov (https://www.alzheimers.gov/clinical-trials/find-clinical-trials) or the Alzheimer's Association (https://www.alz.org/alzheimers-dementia/research-and-progress/clinical-trials) continue to administer and give access to clinical trials for Alzheimer's and dementia research. Source:

- Source: Mainly Google AI and the other articles linked above in the other bullet points

Data-Set:

MetaData Data-Set

- MetaData is the data set that describes who the patient is and their medical/neuropathological background.
  - How it was collected:
    - Brains were collected using highly optimized brain preparation methods
    - Donors were included if death occurred within the specific time of data collection
      - 84 donors
      - Age: minimum 65 years (mean of data set was 88 years)
      - Sex: both sexes included (51 females, 33 males)
  - Overall Content:
    - Column Descriptions:
      - Demographic info (age, sex, race), study source (ACT/ADRC), pathology measures (Braak stage, CERAD score, Thal phase), comorbidities (Lewy body, LATE, microinfarcts, atherosclerosis, arteriolosclerosis), molecular quality metrics (RIN, brain pH), and whether donors were “severely affected" or not.
    - Row Descriptions:
      - Each row represents a unique donor.
  - Techniques
    - Donor Recruitment: Patients were recruited through the Adult Changes in Thought (ACT) study and the University of Washington Alzheimer’s Disease Research Center (ADRC).
    - Brain collection: Postmortem brain samples were collected; tissue was processed for histology, RNA/DNA extraction, and sequencing.
    - Neuropathology Assessment:
      - Immunohistochemistry (IHC): Used to stain brain sections for hallmark Alzheimer’s pathologies
      - Used staging systems
      - Quantitative Neuropathology: Counts of plaques, tangles, neuronal density, and comorbidities
      - Molecular Quality Control: RNA integrity number (RIN) and brain pH were measured to ensure sample quality for sequencing.
  - Pathological/Physiological Processes:
    - Includes both physiological and pathological processes including demographic risk factors, cognitive decline, and neuropathological information such as amyloid plaques, Tau tangles, vascular injury, and co-morbid neurodegenerative changes.
  - Units

        - Age (and age at death): in years
        - Categorical: sex, race, etc
        - Quantitative (reported as integers): RIN, cognitive scores, and counts of microinfarcts
    - Source: https://www.nature.com/articles/s41593-024-01774-5 ; csv file: UpdatedMetaData.csv

Luminex Data-Set

- LuminexData is the data set that measures protein biomarkers in the brain tissue (amyloid-beta and tau)
    - How it was collected:
        - Brain tissue samples from 84 postmortem donors, the same ones studied in the MetaData data-set, were processed to extract proteins.
            - Protein lysates were then analyzed using bead-based immunoassays on the Luminex platform, which allows multiplexed measurement of Alzheimer's disease biomarkers.
            - Each donor corresponds to one record in the dataset, matched by Donor ID.
    - Overall content:
        - Column Descriptions:
            - The dataset contains quantitative values for four core Alzheimer's disease biomarkers:
                - Aβ40 (amyloid-beta 40)
                - Aβ42 (amyloid-beta 42)
                - tTau (total Tau)
                - pTau (phosphorylated Tau)
        - Row Descriptions:
            - Each row represents a unique donor (same as MetaData Data-Set)
    - Techniques
        - Multiplex Bead-Based Immunoassay: Antibody-coated beads bind to specific proteins in brain lysates.
        - Luminex xMAP Technology: Each bead set carries a unique fluorescent signature, and bound proteins are detected via reporter fluorescence.
        - Fluorescence Detection: A dual-laser flow cytometry system reads bead signatures and reporter intensity, converting them into quantitative values via calibration curves.
    - Pathological/Physiological Processes:
        - Describes the pathological protein aggregation processes of Alzheimer's disease, specifically amyloid-beta deposition (Aβ40 and Aβ42) and Tau pathology (total Tau and phosphorylated Tau).
    - Units

        - All biomarker concentrations are expressed in picograms per microgram (pg/μg) of total protein.
    - Source: https://www.nature.com/articles/s41593-024-01774-5 ; csv file: UpdatedLuminex.csv

Original Data Table

- All data comes from the 2024 journal article "Integrated multimodal cell atlas of Alzheimer's disease"
    - Link to article: https://www.nature.com/articles/s41593-024-01774-5#peer-review
        - The full, unfiltered data-sets can be downloaded through the "Supplementary Tables 1-8" link under the Supplementary information section of the journal article.
    - Bias:
        - The 108 authors of the article where the data-sets were derived from "declare no competing interests" underneath the Ethics declaration section. According to *Nature Neuroscience*, the work has also been peer reviewed anonymously. Therefore, we found **no signs of any significant, identifiable bias** that may have influenced the data-sets.
    - Assumptions and Limitations:
        - It's assumed that all the donors' demographic information recorded in the MetaData data-set is true and that all procedures in collecting the data were performed accurately and honestly.
        - The racial diversity of the selected cohort is very limited. After surveying the group's individual races, 81 of the 84 postmortem donors checked their race as "White" with only 3 of the 81 also specifying themselves as being "Mixed" race; the other 3 non-white donors were checked as "Asian".

Data Analysis:

After browsing the two given data-sets, we began brainstorming some questions that can be answered by analyzing the data, eventually becoming interested in betaamyloid values. We ultimately chose to explore the primary question "Does overall AD neuropathological change (OADC) have a higher correlation with betaamyloid-40 (Aβ40) or betaamyloid-42 (Aβ42) concentration in American Alzheimer's patients?" and the secondary, more supplementary question "Are Aβ40 and Aβ42 concentrations correlated with each other?"

Answering the secondary question entails a linear regression analysis with a scatterplot to determine the betaamyloid ratio correlation. On a more difficult end, answering our primary question requires sorting and graphing data from both data-sets, MetaData's OADC categories as the independent variable and both Betaamyloid-40 and 42 concentrations from Luminex as separate dependent variables, and then performing two paired, one-tailed, one-way ANOVA tests separately on the betaamyloid concentrations, one ANOVA test for

Aβ42 and another for Aβ40, across the progressive OADC stages. If one or both groups pass the ANOVA, we'll run a Bonferroni post-hoc test to observe which OADC groups have significant betaamyloid concentration change across them.

Here's an outline of the steps we went through to develop the following code:

- Step 1: Make "Patient" object that combines values from MetaData and Luminex data-sets
- Step 2: Create scatter plot marking patients' Aβ40 and Aβ42 concentrations
- Step 3: Calculate linear regression and R^2 value of the scatter plot
- Step 4: Organize Aβ40 and Aβ42 values into sub-lists based on patients' OADC
- Step 5: Graph OADC's effect on Aβ concentrations on bar graph (to show means) and violin plot (to show general distribution)
- Step 6: Run paired, one-tailed, one-way ANOVA tests separately on Aβ40 and Aβ42 data
- Step 7: Perform Bonferroni post-hoc test(s) if passed ANOVA

In [1]:
```
#1) Make "Patient" object that combines values from MetaData and Luminex .csv data-

import csv
import warnings
import re

class Patient:

    all_patients = []
    death_age = []
    education_lvl = {}

    # Setting Patient object constructor
    def __init__(self, DonorID, ABeta40: float, ABeta42: float, tTau: float, pTau:
        self.DonorID = DonorID
        self.ABeta40 = ABeta40
        self.ABeta42 = ABeta42
        self.tTau = tTau
        self.pTau = pTau
        self.sex = None
        self.death_age = None
        self.ed_lvl = None
        self.cog_stat = None
        self.age_symp_on = None
        self.age_diag = None
        self.head_inj = None
        self.thal_score = None
        self.oadc = None  # Overall AD neuropathological Change (Not AD / Low / Int
        Patient.all_patients.append(self)

    # Defining some getter methods (most of them unused / already there from templa
    def __repr__(self):
        return f"{self.DonorID} | sex: {self.sex} | ABeta40 {self.ABeta40} | tTau {
    def get_id(self):
        return self.DonorID
    def get_ABeta40(self):
```

```
        return self.ABeta40
    def get_ABeta42(self):
        return self.ABeta42
    def get_thal(self):
        return self.thal_score
    def get_death_age(self):
        return self.death_age


    @classmethod
    def combine_data(cls, filename: str):
        # Read metadata
        with open(filename, encoding="utf8") as f:
            reader = csv.DictReader(f)
            meta_rows = list(reader)
            fieldnames = reader.fieldnames or []

        # Build a fast lookup by Donor ID
        meta_by_id = {}
        for r in meta_rows:
            did = (r.get("Donor ID") or "").strip()
            if did:
                meta_by_id[did] = r

        # Identify the OADC/ADNC column, robust to naming variants
        oadc_col = None
        candidates = []
        for fn in fieldnames:
            low = fn.lower()
            if (
                "oadc" in low
                or "adnc" in low
                or (
                    "overall" in low
                    and ("ad" in low or "alzheimer" in low)
                    and ("neuro" in low or "neuropath" in low)
                    and ("change" in low or "nc" in low)
                )
            ):
                candidates.append(fn)
        if candidates:
            # Prefer the shortest, most specific header
            oadc_col = sorted(candidates, key=len)[0]

        def _norm_oadc_val(v):
            if v is None:
                return None
            s = str(v).strip()
            if not s or s.lower() in {"na", "nan", "none"}:
                return None
            sl = s.lower()

            # Try numeric codes first (0–3 sometimes stored as "0", "1.0", etc.)
            try:
                n = int(float(s))
                mapping = {0: "Not AD", 1: "Low", 2: "Intermediate", 3: "High"}
```

```
                if n in mapping:
                    return mapping[n]
            except Exception:
                pass

            # Keyword mapping (also catches forms like "ADNC-High", "OADC: low", et
            if "not" in sl and "ad" in sl:
                return "Not AD"
            if "low" in sl:
                return "Low"
            if "inter" in sl:
                return "Intermediate"
            if "high" in sl:
                return "High"

            # Fallback: title-case whatever is there
            return s.title()

        # Update instantiated patients
        for p in Patient.all_patients:
            r = meta_by_id.get(p.DonorID)
            if not r:
                warnings.warn(f"No metadata found for Donor ID {p.DonorID}")
                continue

            if r.get("Sex"):
                p.sex = r["Sex"]

            if r.get("Age at Death"):
                try:
                    p.death_age = int(r["Age at Death"])
                except ValueError:
                    p.death_age = None

            if r.get("Highest level of education"):
                p.ed_lvl = r["Highest level of education"]

            if r.get("Cognitive Status"):
                p.cog_stat = r["Cognitive Status"]

            if r.get("Age of onset cognitive symptoms"):
                try:
                    p.age_symp_on = int(r["Age of onset cognitive symptoms"])
                except ValueError:
                    p.age_symp_on = None

            if r.get("Age of Dementia diagnosis"):
                try:
                    p.age_diag = int(r["Age of Dementia diagnosis"])
                except ValueError:
                    p.age_diag = None

            if r.get("Known head injury"):
                p.head_inj = r["Known head injury"]

            if r.get("Thal"):
```

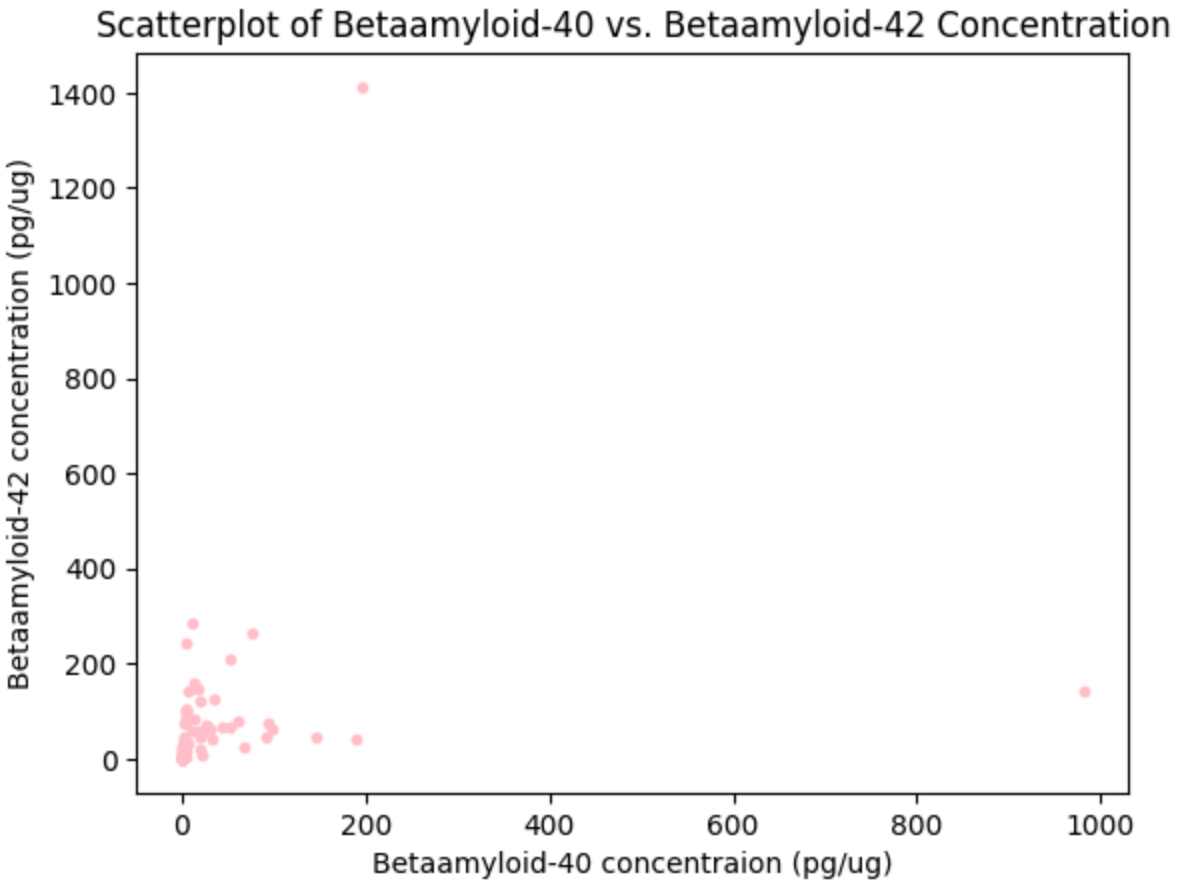
```
                m = re.search(r"(\d+)", r["Thal"])
                p.thal_score = int(m.group(1)) if m else None

            # Robust OADC extraction/normalization
            raw_oadc = r.get(oadc_col) if oadc_col else None
            p.oadc = _norm_oadc_val(raw_oadc)

    @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'] or '').strip(),
                    ABeta40 = float(row['ABeta40 pg/ug']),
                    ABeta42 = float(row['ABeta42 pg/ug']),
                    tTau    = float(row['tTAU pg/ug']),
                    pTau    = float(row['pTAU pg/ug'])
                )

            # sort by Donor ID so printing is stable; metadata join does not rely o
            Patient.all_patients.sort(key = Patient.get_id)
            Patient.combine_data(other_file)
```

In [2]:

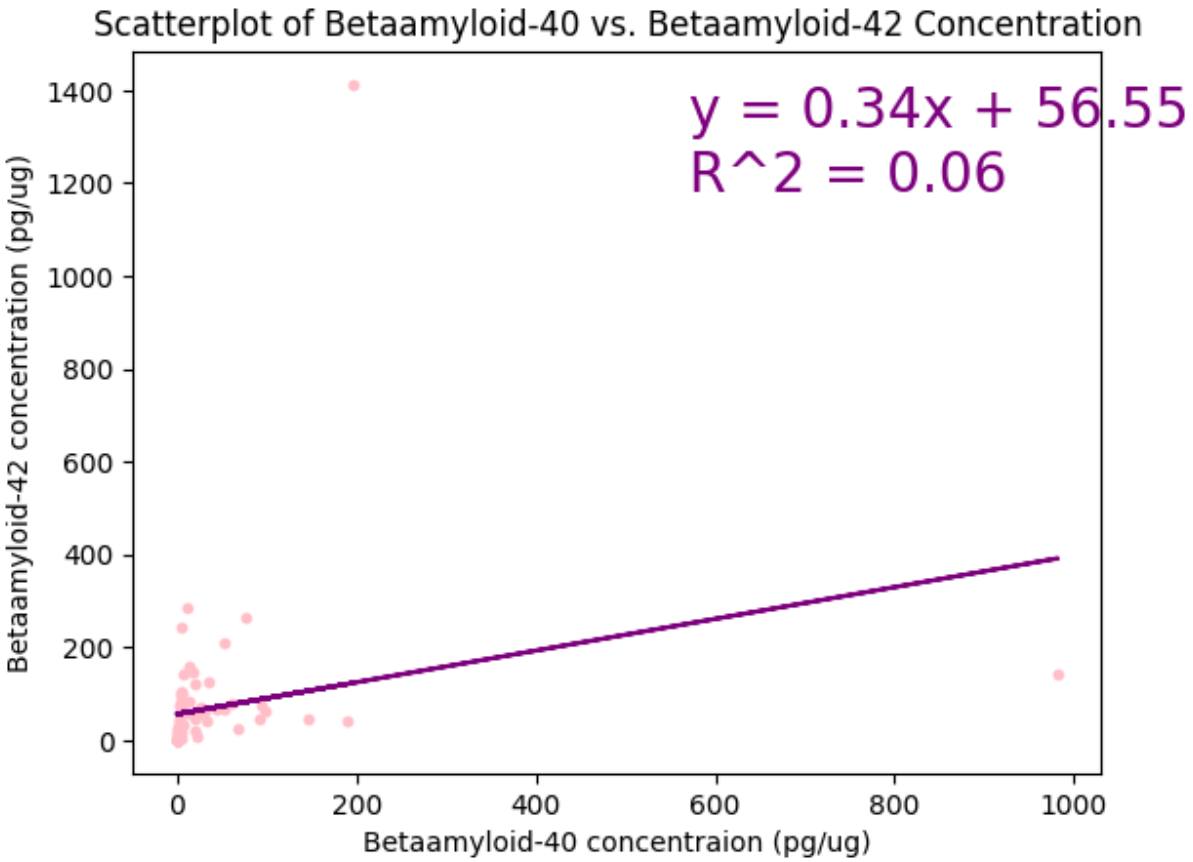
```
#2) Create scatter plot marking patients' Aβ40 and Aβ42 concentrations

from patient import Patient

from matplotlib import pyplot as plt
import numpy as np
from sklearn.linear_model import LinearRegression

# Declare lists for storing all donors' bA40 and bA42 concentrations
bA40_scores = []
bA42_scores = []

# Combine data-sets in Patient class
Patient.instantiate_from_csv('UpdatedLuminex.csv', 'UpdatedMetaData.csv')

# Adding each patient's betaamyloid scores to the lists above
for p in Patient.all_patients:
    bA40_scores.append(p.ABeta40)
    bA42_scores.append(p.ABeta42)

print(bA40_scores)
print(bA42_scores)


# Old code using empty, external csv file to input and retrieve scatterplot data
#
# Putting list data into DataFrame and writing it to external csv file
# bAData = pd.DataFrame({
```

```
#     "bA40 Concentration": bA40_scores,
#     "bA42 Concentration": bA42_scores
# })
# bAData.to_csv("patient_data.csv", index=False)
# print("Data sent to 'patient_data.csv' file")
#
# scatterData = pd.read_csv("patient_data.csv")
# x = scatterData["bA40 Concentration"].values.reshape(-1, 1)
# y = scatterData["bA42 Concentration"].values
#

# Set data values to x- and y-axes
x = np.array(bA40_scores).reshape(-1, 1)
y = np.array(bA42_scores)

# Design and display the scatterplot using matplot
plt.scatter(x, y, s=10, color="pink")
plt.xlabel("Betaamyloid-40 concentraion (pg/ug)")
plt.ylabel("Betaamyloid-42 concentration (pg/ug)")
plt.title("Scatterplot of Betaamyloid-40 vs. Betaamyloid-42 Concentration")
plt.show()
```

```
[0.019621053, 0.215789474, 0.000597895, 60.76631579, 5.136842105, 3.991578947, 11.84
526316, 2.529473684, 1.127368421, 0.526168105, 1.944210526, 2.671578947, 52.6421052
6, 196.732, 1.718947368, 145.2547368, 5.095789474, 3.532631579, 31.56526316, 1.84315
7895, 1.127368421, 1.633684211, 17.22105263, 2.004210526, 91.74842105, 20.21157895,
4.794736842, 0.030147368, 3.594736842, 53.01263158, 5.176842105, 2.062105263, 1.4126
31579, 5.522105263, 97.8, 0.007088421, 981.444, 25.29578947, 0.000882947, 93.6768421
1, 0.000804526, 0.001155368, 1.655789474, 12.87684211, 11.41894737, 18.994, 189.2905
263, 66.77578947, 0.000688421, 0.215789474, 43.23368421, 4.864210526, 0.661052632,
0.009426737, 1.412631579, 0.263157895, 0.001077758, 1.547368421, 0.001261053, 0.0001
30411, 0.000597684, 21.20947368, 76.91789474, 31.71157895, 0.072506632, 17.82736842,
1.827368421, 34.42947368, 1.375789474, 3.967368421, 12.15368421, 7.491578947, 5.3021
05263, 0.036216842, 0.079678842, 1.450526316, 0.065191789, 5.010526316, 20.53894737,
1.593684211, 7.130526316, 21.42315789, 2.421052632, 0.000981053]
[0.971578947, 2.744210526, 0.147157895, 80.26631579, 16.15684211, 101.8305263, 60.51
157895, 47.70947368, 24.78105263, 16.13789474, 27.60947368, 21.27368421, 209.434736
8, 1412.566961, 28.81368421, 45.72842105, 105.1042105, 95.79263158, 63.37473684, 29.
95578947, 18.94736842, 28.85473684, 58.26631579, 42.51368421, 44.25684211, 123.36842
11, 4.96, 0.525263158, 102.4557895, 67.65473684, 81.13789474, 27.33473684, 12.697894
74, 242.5863158, 60.95368421, 0.245263158, 142.778, 69.98842105, 0.405263158, 74.776
84211, 0.405263158, 0.670526316, 6.554736842, 82.97263158, 287.412, 18.994, 40.19894
737, 24.63789474, 0.137347368, 3.502105263, 68.36631579, 35.27578947, 10.09578947,
0.525263158, 20.18210526, 10.98842105, 0.019621053, 38.15894737, 2.672631579, 7.6663
15789, 0.114736842, 8.842105263, 263.5368421, 39.83789474, 0.204385895, 146.8621053,
18.54736842, 124.4347368, 6.777894737, 31.76315789, 161.0947368, 143.4642105, 75.789
47368, 0.449649126, 0.122631579, 76.22631579, 0.490526316, 88.16947368, 47.93263158,
29.89263158, 33.63789474, 53.87894737, 19.19578947, 0.049052632]
```



In [3]:
```
#3) Calculate linear regression and R^2 value of the AB40 vs. AB42 scatter plot

# Create linear regression model from csv data
model = LinearRegression()
model.fit(x, y)
# Assigning linear regression model values to scatterplot equation
slope = model.coef_[0]
intercept = model.intercept_
r2 = model.score(x, y)
equation = f"y = {slope:.2f}x + {intercept:.2f}\nR^2 = {r2:.2f}"

# Design and display the scatterplot using matplot
plt.scatter(x, y, s=10, color="pink")
plt.plot(x, model.predict(x), color="purple")
plt.xlabel("Betaamyloid-40 concentraion (pg/ug)")
plt.ylabel("Betaamyloid-42 concentration (pg/ug)")
plt.title("Scatterplot of Betaamyloid-40 vs. Betaamyloid-42 Concentration")
plt.text(570, y.max(), equation, color="purple", fontsize=20, verticalalignment="to
plt.show()
```



In [4]:

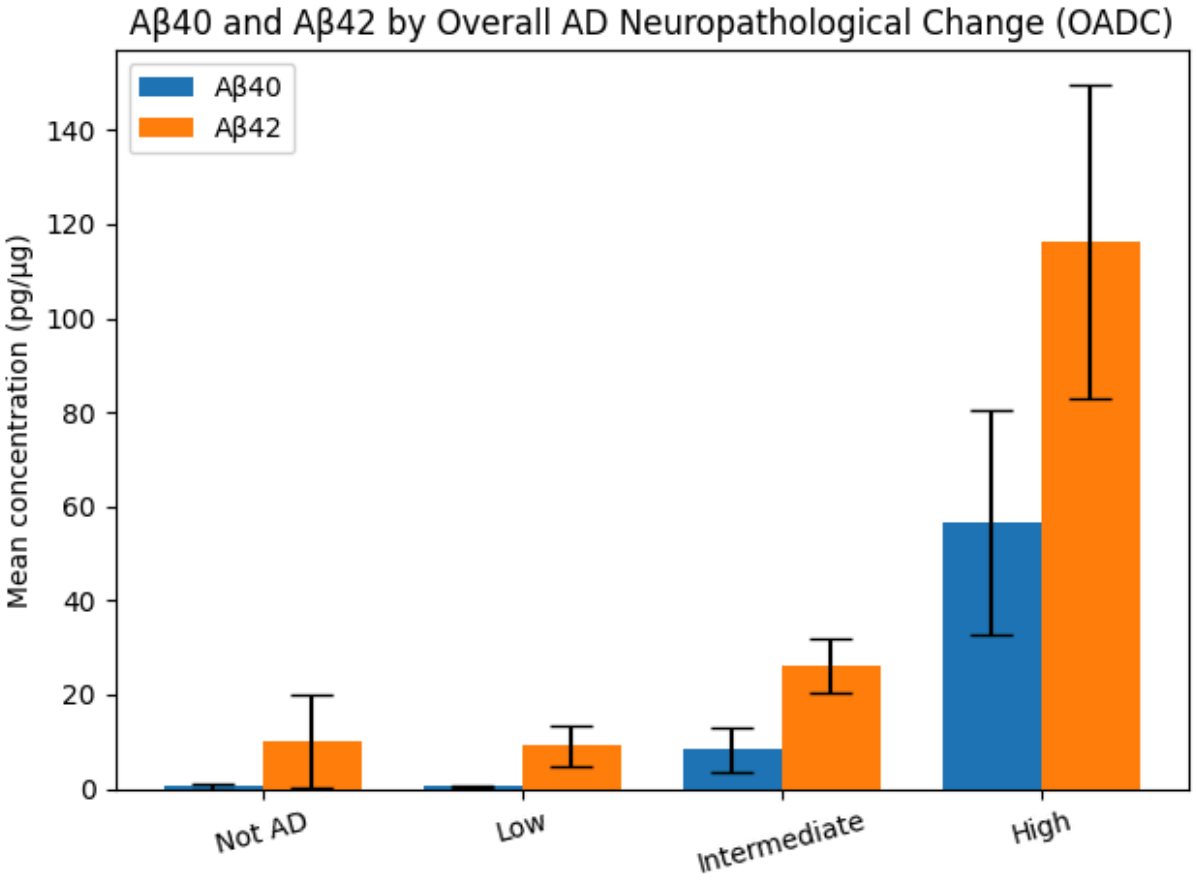
```python
#4) Organize Aβ40 and Aβ42 values into sub-lists based on patients' OADC

import math

# Initialize list of OADC stages
CATS = ["Not AD", "Low", "Intermediate", "High"]

# Normalize OADC labels into consistent categories
def _norm_oadc(val):
    if val is None:
        return None
    s = str(val).strip().lower()
    if not s or s in ("nan", "none"):
        return None
    if "not" in s and "ad" in s:
        return "Not AD"
    if "low" in s:
        return "Low"
    if "inter" in s:
        return "Intermediate"
    if "high" in s:
        return "High"
    try:
        m = int(s)
        return ["Not AD", "Low", "Intermediate", "High"][m]
    except Exception:
        return s.title()
# Collect values of a given metric (Aβ40 or Aβ42) for one category
def _vals(metric, cat):
```

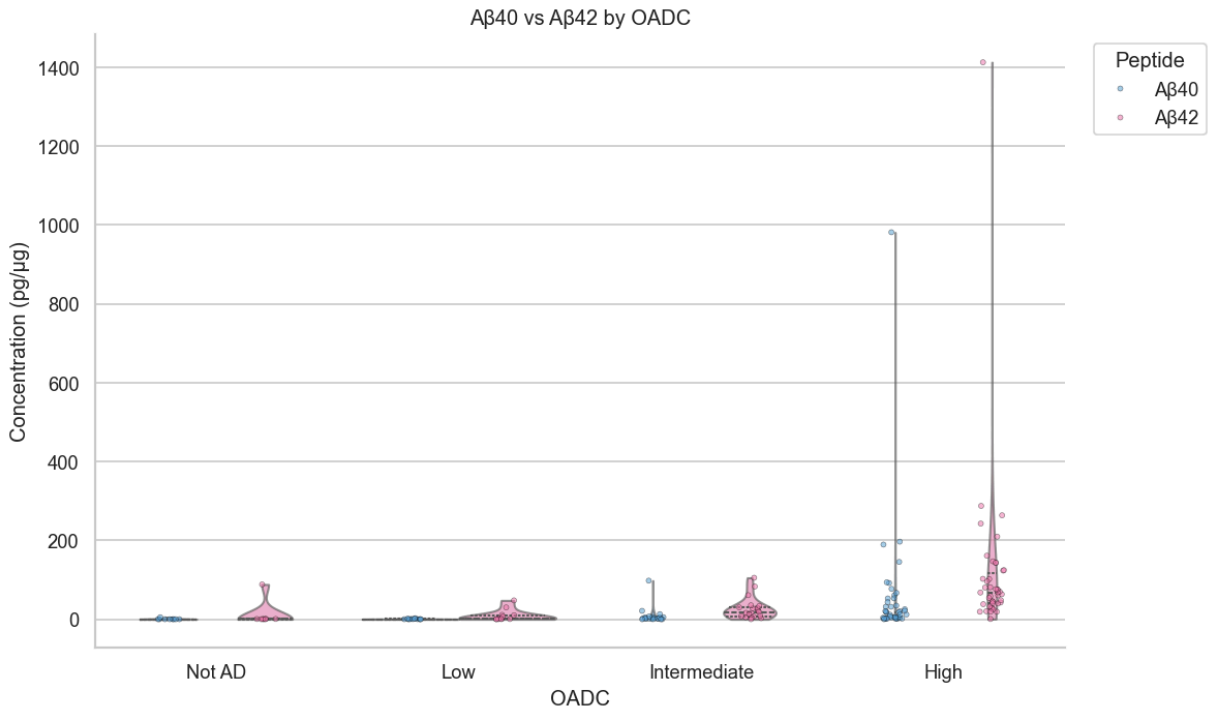
```
    vals = []
    for p in Patient.all_patients:
        key = _norm_oadc(getattr(p, "oadc", None))
        if key == cat:
            v = getattr(p, metric, None)
            if v is not None and not (isinstance(v, float) and math.isnan(v)):
                vals.append(v)
    return vals
# Build sub-lists for each OADC stage for Aβ40 and Aβ42
groups40 = [_vals("ABeta40", c) for c in CATS]
groups42 = [_vals("ABeta42", c) for c in CATS]
```

In [5]:

```
#5) Graph OADC's effect on Aβ concentrations on bar graph (to show means) and violi

import pandas as pd, seaborn as sns, matplotlib.pyplot as plt
import numpy as np
from matplotlib import pyplot as plt

# Make Bar Graph
# Compute mean concentration for each OADC group
means40 = [np.mean(g) if len(g) else np.nan for g in groups40]
means42 = [np.mean(g) if len(g) else np.nan for g in groups42]

# Compute standard deviations for each OADC group
se40 = [np.std(g, ddof=1)/np.sqrt(len(g)) if len(g) >= 2 else np.nan for g in group
se42 = [np.std(g, ddof=1)/np.sqrt(len(g)) if len(g) >= 2 else np.nan for g in group

x = np.arange(len(CATS))
w = 0.38

# Plot grouped bar chart with error bars
fig, ax = plt.subplots()
ax.bar(x - w/2, means40, w, label="Aβ40", yerr=se40, capsize=8)
ax.bar(x + w/2, means42, w, label="Aβ42", yerr=se42, capsize=8)

ax.set_xticks(x)
ax.set_xticklabels(CATS, rotation=15)
ax.set_ylabel("Mean concentration (pg/µg)")
ax.set_title("Aβ40 and Aβ42 by Overall AD Neuropathological Change (OADC)")
ax.legend()

plt.tight_layout()
plt.show()

# Making Violin Plot
# Build long-form DataFrame for plotting
rows = []
for p in Patient.all_patients:
    if getattr(p, "oadc", None) is None:
        continue
    rows += [
        {"OADC": p.oadc, "Peptide": "Aβ40", "Concentration": p.ABeta40},
        {"OADC": p.oadc, "Peptide": "Aβ42", "Concentration": p.ABeta42},
    ]

df = pd.DataFrame(rows).dropna(subset=["OADC","Concentration"])
```

```
order   = ["Not AD","Low","Intermediate","High"]
palette = {"Aβ40":"#5DA5DA","Aβ42":"#F17CB0"}
sns.set(style="whitegrid", rc={"axes.spines.right":False, "axes.spines.top":False})

# Create violin plot with overlaid strip plot
plt.figure(figsize=(10,6), dpi=120)
ax = sns.violinplot(data=df, x="OADC", y="Concentration", hue="Peptide",
                    order=order, palette=palette, inner="quart",
                    split=False, cut=0, dodge=True, alpha=0.65)
sns.stripplot(data=df, x="OADC", y="Concentration", hue="Peptide",
              order=order, palette=palette, dodge=True, jitter=True,
              alpha=0.55, size=3, linewidth=0.3, edgecolor="black")

handles, labels = ax.get_legend_handles_labels()
by_label = dict(zip(labels, handles))
ax.legend(by_label.values(), by_label.keys(), title="Peptide", bbox_to_anchor=(1.02

# Set axis labels and title
ax.set_title("Aβ40 vs Aβ42 by OADC")
ax.set_ylabel("Concentration (pg/µg)")
plt.tight_layout(); plt.show()
```





In [6]:
```python
#6) Run paired, one-tailed, one-way ANOVA tests separately on Aβ40 and Aβ42 data

import numpy as np
from scipy import stats
from itertools import combinations

# Helper function to run one-way ANOVA
def _anova(groups):
    nonempty = [g for g in groups if len(g) > 0]
    if len(nonempty) >= 2:
        F, p = stats.f_oneway(*nonempty)
        return F, p
    return np.nan, np.nan
# Run ANOVA for Aβ40 and Aβ42 groups
F40, p40 = _anova(groups40)
F42, p42 = _anova(groups42)
# Helper to convert p-values into significance strings
def _sig_str(p):
    if not (p == p):
        return "n/a (not enough data)"
    # Report significance based on threshold
    return "ns (p > 0.05)" if p > 0.05 else "significant (p ≤ 0.05)"
# Print results of ANOVA tests with significance status
print(f"Aβ40 ANOVA across OADC: p = {p40:.4g} → {_sig_str(p40)}")
print(f"Aβ42 ANOVA across OADC: p = {p42:.4g} → {_sig_str(p42)}")
```

```
Aβ40 ANOVA across OADC: p = 0.2015 → ns (p > 0.05)
Aβ42 ANOVA across OADC: p = 0.0432 → significant (p ≤ 0.05)
```

In [ ]:
```python
#7) Perform Bonferroni post-hoc test for Aβ42 (since it passed ANOVA)

# Bonferroni post hoc test conduction
def bonferroni_posthoc(groups, labels, name):
    # Generate all unique pairwise comparisons (i, j) across groups
```

```
    pairs = [(i, j) for i, j in combinations(range(len(groups)), 2)
             if len(groups[i]) >= 2 and len(groups[j]) >= 2]
    if not pairs:
        return
    m = len(pairs)
    print(f"\n{name} Bonferroni p-values:")
    # Loop through each pair of groups
    for i, j in pairs:
        _, p = stats.ttest_ind(groups[i], groups[j], equal_var=False, nan_policy='o
         # Apply Bonferroni correction by multiplying p-value by number of comparis
        adj = min(p*m, 1.0)
        print(f"  {labels[i]} vs {labels[j]}: \n  Bonferroni p = {adj:.4g} → {_sig_
# Run Bonferroni post hoc test for Aß42 groups
bonferroni_posthoc(groups42, CATS, "Aβ42")
```

```
Aβ42 Bonferroni p-values:
  Not AD vs Low:
  Bonferroni p = 1 → ns (p > 0.05)

  Not AD vs Intermediate:
  Bonferroni p = 1 → ns (p > 0.05)

  Not AD vs High:
  Bonferroni p = 0.02235 → significant (p ≤ 0.05)

  Low vs Intermediate:
  Bonferroni p = 0.1604 → ns (p > 0.05)

  Low vs High:
  Bonferroni p = 0.01625 → significant (p ≤ 0.05)

  Intermediate vs High:
  Bonferroni p = 0.06499 → ns (p > 0.05)
```

Verify and validate your analysis:

Verification:

- **Low Aβ42-Aβ40 Correlation**: Although we were surprised by the lower overall Aβ42:Aβ40 ratio/slope calculated by the linear regression, we didn't pay that much attention to it given how the R^2 value of "Aβ42 vs. Aβ40 concentrations" came out as a low 0.06, meaning that only 6% of the variation of Aβ42 concentration is explained by Aβ40 concentration. The low R^2 value told us that there wasn't much of a consistent ratio between Aβ42 and Aβ40 among the donors. This answer made sense to us given how different each donor from data-set was; every patient had their own combination of OADC stage, gender, education level, Tau levels, and other characteristics that can alter betaamyloid values, likely making the Aβ42:Aβ40 ratio different for each person.
- **Significant Correlation between OADC and Aβ42**: As shown by our graphs as well as ANOVA and Bonferroni post-hoc tests, betaamyloid-42 concentration significantly increases between low OADC stages (*No AD* and *Low*) and *High* OADC. This result,

showing that Aβ42 can correlate to AD neuropathological change, made sense to us. Betaamyloid is a key protein associated with causing and progressing Alzheimer's Disease, and we knew beforehand that OADC stages described in MetaData was partially rated by tests measuring to betaamyloid, as briefly mentioned in the study where the data-sets came from; the notion that betamyloid spurs AD neuropathological change. While we didn't know exactly why Aβ40 didn't show a significant correlation to OADC like Aβ42, we did find out that Aβ42 was generally perceived as more toxic during our background research, suggesting Aβ40 may not be potent enough to significantly influence OADC in these data-sets of donors.

Validation:

- **Lei Gu and Zhefeng Gao's "Alzheimer's Aβ42 and Aβ40 peptides form interlaced amyloid fibrils"**: This article from the *Journal of neurochemistry* studies the efficiency of Aβ42 and Aβ40 in aggregation by observing them form amyloid fibrils. Through using electron paramagnetic resonance (EPR) spectroscopy to observe fibril growth, the researchers found that "Aβ40 is not as efficient as Aβ42 in terms of being incorporated into Aβ42 fibrils" and later concluded that populations of solely Aβ42 were responsible for "amyloid pathology in AD," in part due to Aβ40 aggregating too slowly. These conclusions not only validate our analysis finding that the significant correlation between overall AD neuropathological change and Aβ42, but it also implies that Aβ42 and Aβ40 grow at different rates, suggesting the Aβ42:Aβ40 within AD patients is not constant and naturally changes over time.
    - **Citation**: Gu, L., & Guo, Z. (2013). Alzheimer's Aβ42 and Aβ40 peptides form interlaced amyloid fibrils. Journal of neurochemistry, 126(3), 305–311. https://doi.org/10.1111/jnc.12202
- **Pérez-Grijalba, et al.'s "Plasma Aβ42/40 Ratio Detects Early Stages of Alzheimer's Disease and Correlates with CSF and Neuroimaging Biomarkers in the AB255 Study"**: This published article, also from the *Journal of neurochemistry*, focuses on the Aβ42:Aβ40 ratio itself and how a low ratio may be a signifier for mild cognitive impairment and early Alzheimer's Disease. The idea that a certain Aβ42:Aβ40 ratio can be a marker for certain AD progression stages validates our linear regression finding that the Aβ42 concentration doesn't correlate with Aβ40 concentration well (in other words, Aβ42:Aβ40 ratio isn't constant among the donors in the data-sets), given that it wouldn't make sense to consider Aβ42:Aβ40 ratio as any signifier at all if it always remained the same in each person. Moreover, some of the graphs and results in the article display Aβ42:Aβ40 ratio changing predictably over different independent variables, like Aβ-PET and CSF biomarkers.
    - **Citation**: Pérez-Grijalba, V., Romero, J., Pesini, P., Sarasa, L., Monleón, I., San-José, I., Arbizu, J., Martínez-Lage, P., Munuera, J., Ruiz, A., Tárraga, L., Boada, M., & Sarasa, M. (2019). Plasma Aβ42/40 Ratio Detects Early Stages of Alzheimer's Disease and Correlates with CSF and Neuroimaging Biomarkers in the AB255 Study. The journal

of prevention of Alzheimer's disease, 6(1), 34–41. https://doi.org/10.14283/jpad.2018.41

- **National Institute of Aging's "What Happens to the Brain in Alzheimer's Disease?"**: While this source is better for providing background information on AD than validating results, the article section about Amyloid plaques does note how "The beta-amyloid 42 form is thought to be especially toxic" as it aggregates and interrupts cell communication, giving complementary reasoning that may explain why we found that Aβ42 concentration significantly increases as OADC progresses through stages of lower to high AD cognitive change.
  - **Citation**: "What Happens to the Brain in Alzheimer's Disease?" National Institute on Aging, 19 Jan. 2024, https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease

Conclusions and Ethical Implications:

Conclusions:

- For our first question: "Are betaamyloid-40 and betaamyloid-42 correlated with each other?"
  - We concluded that Aβ42:Aβ40 ratio not correlated, suggesting it changes amongst different patients. This is proven through the linear regression plot as the R^2 value was 0.06, proving that there is little, to no correlation between the two beta-amyloid concentrations.
- For our second question: "Does overall AD neuropathological change (OADC) have a higher correlation with betaamyloid-40 or betaamyloid-42 concentration in American Alzheimer's patients?"
  - We concluded that OADC has high correlation to Aβ42, but not Aβ40. Aβ42 concentration at lower OADC groups is significantly different than the High OADC. This is proven through the one-way ANOVA and post-doc Bonferroni test. From the ANOVA, Aβ40 concentration means were found to be not statistically different across OADC groups with a p-value of 0.2015; however, Aβ42 concentrations were found to be statistically different with a p-value of 0.0432. Since Aβ42 passed the ANOVA test, we performed a post doc Bonferroni test where we found that concentrations within the Not AD and Low OADC groups were significantly different from concentrations within the High OADC group. Our significance level (alpha) across all the tests was 0.05.

Ethical Implications:

- One ethical implication is that people may perceive Aβ42 concentration as dementia-inducing, causing anxiety. This perception might lead to emotional distress in individuals who undergo testing and find higher Aβ42 levels, even if they are asymptomatic or their

risk of developing dementia is not definitively established. This may also spark fear if the individual was just diagnosed with Alzheimer's and they are known to have higher concentrations of Aβ42. Such anxiety can influence life choices, insurance decisions, and doctor–patient relationships and may stigmatize individuals based solely on biomarker results rather than the full clinical picture. This makes it important for physicians & hospitals to communicate these findings responsibly, ensuring that patients understand the difference between correlation and causation, as well as the limitations of current biomarker research.

- A second ethical implication is that reliable biomarkers, such as Aβ42, can guide earlier AD detection and resource planning. Earlier detection could enable patients and families to make informed decisions about long-term care, financial planning, and participation in clinical trials, while also helping physicians initiate supportive interventions sooner. From a healthcare system perspective, validated biomarkers could help hospitals and policymakers allocate resources more efficiently, preparing for the anticipated rise in AD cases as populations age. However, this must be balanced against overdiagnosis risk and unfair access to testing. Overdiagnosis could occur if individuals are labeled as at risk for Alzheimer's disease based solely on biomarker levels, even when they may never progress to clinical dementia. Such premature labeling may expose patients to stigma, psychological harm, or unnecessary medical interventions.

Limitations and Future Work:

Limitations:

- One limitation is that honing in only on beta-amyloid-40 and beta-amyloid-42 may oversimplify the complex biological processes underlying Alzheimer's disease and overall Alzheimer's Disease Neuropathologic Change (OADC). While beta-amyloid accumulation is a well-established contributor, it is only one piece of a much broader pathological set of causations. Other critical contributors, such as tau pathology, neuroinflammation, vascular changes, and synaptic dysfunction, also play significant roles in disease onset and progression. By excluding these factors, the project risks presenting an incomplete picture of AD pathology, potentially limiting the depth and applicability of the conclusions. This narrow focus may also restrict the ability to identify interactions between biomarkers, which could be essential for understanding why some patients progress more rapidly than others despite similar betaamyloid concentrations.
- A second limitation of our study is the relatively small sample size (n=84), particularly when broken down across the different OADC groups. With only 84 total donors, the number of individuals represented within each neuropathological category is limited, and this uneven distribution makes it challenging to draw strong statistical conclusions. For example, some OADC groups may have very few cases, which reduces the reliability of comparisons and increases the likelihood that observed differences are due to random variation rather than true biological effects.

Future Work:

- If we were to do this project again, we'd like to explore:
  - How neurological tests relate to Aβ40 and Aβ42 concentrations and OADC.
    - This could be an interesting expansion of our study because cognitive assessments such as MMSE, CASI, and MOCA provide direct clinical measures of memory, attention, and executive function, while the biomarkers capture underlying molecular changes in the brain. Linking these two factors could help clarify whether higher concentrations of Aβ42 or altered Aβ40:Aβ42 ratios translate into measurable differences in cognitive performance across OADC groups.
  - The further examination of Aβ42:Aβ40 ratio rather than exploring beta-amyloid-40 and beta-amyloid-42 separately.
    - This could be an interesting expansion of our study because this ratio has been suggested in prior research to better reflect pathological imbalance in beta-amyloid processing. While concentrations of Aβ40 and Aβ42 separately provide valuable information, the ratio between them may capture subtle shifts in production or clearance that are more strongly associated with plaque formation and disease severity.

NOTES FROM TEAM:

Current Progress:

9/9/25-9/11/25: In class, we did a lot of preliminary research and both caught each other up on how to push and pull files through GitHub. We established our shared files and folders for us to use throughout the project. We exchanged numbers and established a day to complete more research. Specifically, we divided up the work by Logan doing the first 5 bullet points, Luke doing the last 5, and both of us completing the 6th bullet point (or whoever gets to it first). We also brainstormed and selected a question to explore and analyze for the project.

9/14/25: We both completed more research and started to explore the data sets in more depth. We completed the work on our own, but checked in throughout the day to update each other about our progress and any questions we had. Completed 7 total bullet points and extracted the important information from the data sets (both Meta and Luminex).

9/16/25: In class, we started coding and focusing on our question. We went line by line of the code given to us and made sure we both had a common understanding of what was being executed. Once we finished, we started editing the code to fit our project question.

9/17/25: Texted about current project progress and established that we would like to finish all of our research bullet points by 9/18. We also discussed that we would both try to set up

a bar graph fitting our data/question by Thursday's class.

9/18/25: We learned more about data analysis through bar graphs in today's class and started doing more work for the code portions. After graphing out the Tau levels for each education level, however, Logan saw that the Tau score differences were not significant. Since we wanted a research topic that provided a stronger correlation, we switched our research question to "Does betaamyloid 40 or 42 correlate to higher overall AD neuropathological change in patients?" Logan again figured out the Python code and bar graph analysis of our new question and found a much more potent correlation, and Luke is writing this entry still confused on what an ANOVA is.

9/19/25: Finished up all the Disease Background questions and more of the Data Analysis section. We also did some very minor Jupyter polishing to prepare for tonight's Second Jupyter Notebook Check-in submission.

9/24/25: We met out of class during the evening to finalize our statistical data and graphs. We decided to frame our presentation in this way: contextual information using a scatterplot to show IF there is a correlation between amyloid beta 40 and 42 (example: if one decreases, does the other one also decrease?); using a violin plot (and bar graph to add simplicity) to show how concentrations decrease/increase across AD stages; finally, we will prove these findings with a 1 way ANOVA with p values.

9/25/25: We confirmed our statistical tests and graphs with our TA (said they are a great way to go about the presentation). We went through the rubric and started adding the necessary content. Started working on data analysis, conclusions, and other implications (ethical and limitations).

9/29/25: We met outside of class time to wrap up our statistical tests and prepare for our presentation. The main highlight from the meeting was that we realized we needed a post hoc test to analyze beta-amayloid-42 concentrations across the different OADC groups (since it passed ANOVA test). We did this by conducting a Bonferroni post hoc test. We created our presentation and did a quick run-through on our speaking points.

9/30/25: We met an hour before our presentation to go through timing and our "script" together. We practiced 5 times to ensure we were confident in the points we were going to make.

Code History and Sources

Code History

- We used Git and GitHub to work and collaborate on our Jupyter notebook. Here's the link to our GitHub page to look over the commit history of our code: https://github.com/Lukinator3000/AlzheimersModule

Code Sources

- https://chatgpt.com/
- https://medium.com/luteceo-software-chemistry/statistical-analysis-using-f-and-jupyter-notebooks-2e2f31ee4cc1
- https://www.geeksforgeeks.org/data-analysis/data-analysis-and-visualization-with-jupyter-notebook/
- https://medium.com/@SamTaylor92/data-analysis-statistical-analysis-7-11-ce20fd1b1a1a
- https://www.geeksforgeeks.org/python/how-to-perform-a-one-way-anova-in-python/
- https://www.geeksforgeeks.org/r-language/bonferroni-test/

Extra project questions we thought about:

- Does brain pH affect how severe dementia is (in terms of Tau score, demetia diagnosis or age of death)?
- Does presence of dementia correlate to severity in arteriolosclerosis?
- Does betaamyloid 40 or 42 correlate to higher overall AD neoropathological change in patients?
  - 9/18/25 UPDATE: This is now our new research question!
- Does the highest level of education acheived by Alzheimer's patients affect their Tau score, both total and phosphorylated?
  - 9/18/25 UPDATE: Scrapped this question for the one above it.

QUESTIONS FOR TA:

- First check-in:
  - Just to make sure, how should we accurately cite the two datasets already given to us in Canvas?
  - Should we care greatly about the grammar and overall diction of our Jupyter notes, or should we just focus on jotting down concise and readable notes for these sections?
  - How should we cite the sources we used for the Disease Background section, and where should we put the citations?
  - Is it ok if our question studies both tTau and pTau score, or should we only focus on one of them?
  - Answers to questions:
    - "To answer the questions you have for me, you should cite the study the dataset was presented in. The jupyter notebooks do not have to be the most pristine code documentations for the check ins, though I would like to see a polished one at the end. I would like to be able to see what you've done each day (for future reference). For the citations in the disease background, it should be fine to put them as you have been with other sections. Your question is fine to study both pTau and tTau! I think its a great idea to do both if you can."

- Second check-in:
  - How much should we document our code? Should we add more in the data analysis Python code above, or is it sufficient enough?
  - Did we pick the right kind of data analysis method, just to make sure?
  - What recommendations would you suggest we make to improve our code, method, or bar graph display?