Module_1:

Team Members:

Sheza Khan and Alexander Kremsreiter

Project Title:

Nuerodegeneration - Alzheimer's Disease

Project Goal:

This project seeks to answer the questions: How will APOE Genotype correlate to CERAD scores, and how will these factors correlate to a patient's health?

Disease Background:

Fill in information about 11 bullets:

- Prevalence & incidence https://pubmed.ncbi.nlm.nih.gov/38689398/ In the US, prevalence seems to be higher in states with large, diverse populations (ex. California, New York) and roughly in the South. In 2024, approximately 6.9 million people had Alzheimer's disease. In terms of incidence, deaths due to Alzheimer's increased slightly from 2015-2019 to 2020-2021. However, this could be due to outside factors such as the pandemic (ex. hospitals under more stress). In general, deaths are higher in the winter months and dip in the middle of the year. One figure predicts that incidence will increase from 2020 to 2060.
- Economic burden https://pmc.ncbi.nlm.nih.gov/articles/PMC10398271/ Alzheimer's disease increases Medicare cost slightly to patients; a sizeable portion of diseased beneficiaries had a history of Alzheimer's in their claims history. Families and medicaid typically cover a good amount of the cost burden. https://news.northeastern.edu/2024/07/12/new-alzheimers-drug-cost/
- Risk factors (genetic, lifestyle) Alzheimer's disease has a variety of risk factors from genetic (APOE genotype) to external/epigenetic (exposure to things like aluminum metal). This article discusses 20 different risk factors with varying levels of controverys; some, like infectious agents, have been researched more extensively than others (ex. obesity). Factors related to lifestyle/non-biological causes are typically less researched. https://pubmed.ncbi.nlm.nih.gov/31556570/
- Societal determinants Many societal issues impact health, and this could potentially put certain groups at higher risk for developing the disease. For example, those in low-income or disadvantaged environments who aren't able to access health care or higher education. https://www.cdc.gov/alzheimers-dementia/php/sdoh/index.html
- Symptoms Symptoms of Alzheimer's typically consist of cognitive/personality changes, but the specifics within these categories can vary. Interestingly, "preserved skills" such as reading usually are not imapacted early on because they are controlled by parts of the brain affected in later stages of the disease. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447
- Diagnosis Seeing a physician (neurologist) is required for diagnosis. They will look at a few key things mentioned in the symptoms section (ex. memory issues or personality changes) that match symptoms for early stages of the disease. Other labs/brain imaging also might be done. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers/art-20048075
- Standard of care treatments (& reimbursement)
    - No treatment to cure Alzheimer's, all treatments are for early or middle stages. -https://www.nia.nih.gov/health/alzheimers-treatment/how-alzheimers-disease-treated -Galantamine, benzgalantamine, rivastigmine, and donepezil are cholinesterase inhibitors (a group of medicines that block the normal breakdown of acetylcholine, released by motor neurons to activate muscles also has an important role in arousal, attention, learning, memory and motivation.)
        - https://www.drugs.com/drug-class/cholinesterase-inhibitors.html -As Alzheimer's progresses, less acetylcholine is produced, rendinering medications ineffective -Lecanemab and donanemab are medications target the protein beta-amyloid to help reduce amyloid plaques
    - insurance may only cover these medications in specific situations. Medicare Part B covers part of the cost of these medications for patients who meet certain medical criteria. Most health insuracnes have criteria that needs to be met for them to cover the cost of trreatment and care of Alzheimer's patients.
- Disease progression & prognosis (https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-stages/art-20048448)
    - Preclinical: typically only indentified in the research stage, non-noticable symptoms, biomakrers found in blood samples, brain scans, and genetic tests
    - Mild cognitive: mild changes in memory and thinking, wont affect work or relationships, potential memory lapses, diagnosed based on healthcare professional review of symptoms adn biomarkers.
    - Mild dementia: memory loss of recent events, toruble with problem-solving, complex tasks and sound judgement, changes in personality, trouble organizing and expressing thoughts, getting lost or misplacing belongings.
    - Moderate dementia: Show increasingly poor judgment and deepening confusion, experience even greater memory loss, need help with daily activities, and undergo significant changes in personality and behavior.
    - Sever dememntia: lose the ability to communicate, require daily assistance with personal care, expereince a decline in physical abilities.
- Continuum of care providers -https://health.clevelandclinic.org/long-term-care-for-alzheimers-patients Continuum of care options for patients with Alzheimer's disease vary. Some options are more individualized, such as having an in-home caretaker, while others are more communal such as in adult day care centers or assisted living facilities. Some options are more specialized/geared towards those with the disease while others are more broad and the range of care is general to overall lifestyle.
- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology) https://pmc.ncbi.nlm.nih.gov/articles/PMC5290713/ Alzheimer's disease may be attributed to certain physiological phenomenon such as increase in presence of amyloid-beta plaques and neurofibrillary tangles. These affect many areas in the brain, such as the brain stem, thalamus, amygdala, and more. Along with accumulation of plaques/tangles, neuron loss and disruption to neurotransmitter signals are also biological processes associated with the progression of the disease.

- Clinical Trials/next-gen therapies https://www.alzheimers.org.uk/what-we-do/researchers/news/researching-new-drugs-alzheimers-disease In the UK, two new "disease-modifying" drugs are being researched and slowly implemented to treat Alzheimer's disesae. They aim to slow down memory loss by targetting amyloid protein and utilizing the body's immune system to fight/prevent build-up.

Data-Set:

The data set being used is from the Nature Nueroscience group's study on 'Integrated multimodal cell atlas of Alzheimer's disease'. This study's focus was to understand which of the brain's cell types were affected during the progression of Alzheimer's Disease. The focus group of this study: 84 donors (33 men and 51 women) between ages 65-100, with the mean age being 88. The researchers utilized quantitative nueropathy, single nucleus RNA sequencing, and spatial transcriptomics to collect and analyze data. "We collectively profiled 3.4 million high-quality nuclei across all modalities, mapping each to one of 139 molecular cell types" (Gabitto et al.). Additionally, they created a continuous pseudoprogression score (CPS) from quantitative neuropathology, in which donors were ranked along a neuropathological scale to identify molecular and cellular changes.

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

Data Analyis:

In [6]:
```python
import csv
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats

class Patient:
    all_patients = []

    def __init__(self, DonorID, APOE_Genotype, CERAD_score, CognitiveStatus, Age_at_death):
        self.donor_id = DonorID
        self.apoe = APOE_Genotype
        self.CERAD_score = CERAD_score
        self.cognitive_status = CognitiveStatus
        self.age_at_death = Age_at_death
        Patient.all_patients.append(self)

    def __repr__(self):
        return (
            f"Donor ID: {self.donor_id} | "
            f"APOE Genotype: {self.apoe} | "
            f"CERAD score: {self.CERAD_score} | "
            f"Cognitive Status: {self.cognitive_status} | "
            f"Age at death: {self.age_at_death}"
        )

    @classmethod
    def instantiate_from_csv(cls, filename: str):
        # Use utf-8-sig to handle BOM
        with open(filename, encoding="utf-8-sig") as f:
            reader = csv.DictReader(f)

            # Normalize headers: strip spaces and lowercase them
            rows_of_patients = []
            for row in reader:
                normalized_row = {k.strip().lower(): v for k, v in row.items()}
                rows_of_patients.append(normalized_row)

            # Create Patient objects
            for row in rows_of_patients:
                Patient(
                    DonorID=row["donor id"],
                    APOE_Genotype=row["apoe genotype"],
                    CERAD_score=row["cerad score"],
                    CognitiveStatus=row["cognitive status"],
                    Age_at_death=row["age at death"]
                )
```

In [7]:
```python
# 2. Load patients from MetaData.csv and print patients
Patient.instantiate_from_csv("MetaData.csv")
print("Loaded patients:")
for patient in Patient.all_patients:
    print(patient)
```

```
Loaded patients:
Donor ID: H19.33.004 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 80
Donor ID: H20.33.001 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 82
Donor ID: H20.33.002 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 97
Donor ID: H20.33.004 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 86
Donor ID: H20.33.005 | APOE Genotype: 2/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 99
Donor ID: H20.33.008 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 92
Donor ID: H20.33.011 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 93
Donor ID: H20.33.012 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 91
Donor ID: H20.33.013 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 94
Donor ID: H20.33.014 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 82
Donor ID: H20.33.015 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 88
Donor ID: H20.33.016 | APOE Genotype: 2/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 93
Donor ID: H20.33.017 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 69
Donor ID: H20.33.018 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 81
Donor ID: H20.33.019 | APOE Genotype: 3/4 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 87
Donor ID: H20.33.020 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 81
Donor ID: H20.33.024 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 90
Donor ID: H20.33.025 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 94
Donor ID: H20.33.026 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 75
Donor ID: H20.33.027 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 99
Donor ID: H20.33.028 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 94
Donor ID: H20.33.029 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 91
Donor ID: H20.33.030 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 86
Donor ID: H20.33.031 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 87
Donor ID: H20.33.032 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 98
Donor ID: H20.33.033 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 68
Donor ID: H20.33.034 | APOE Genotype: 2/2 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 85
Donor ID: H20.33.035 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 99
Donor ID: H20.33.036 | APOE Genotype: 2/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 100
Donor ID: H20.33.037 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 96
Donor ID: H20.33.038 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 90
Donor ID: H20.33.039 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 96
Donor ID: H20.33.040 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 98
Donor ID: H20.33.041 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 91
Donor ID: H20.33.043 | APOE Genotype: 4/4 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 85
Donor ID: H20.33.044 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 81
Donor ID: H20.33.045 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 77
Donor ID: H20.33.046 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 94
Donor ID: H21.33.001 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: Dementia | Age at death: 80
Donor ID: H21.33.002 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 70
Donor ID: H21.33.003 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 78
Donor ID: H21.33.004 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 93
Donor ID: H21.33.005 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 95
Donor ID: H21.33.006 | APOE Genotype: 3/4 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 97
Donor ID: H21.33.007 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 86
Donor ID: H21.33.008 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 91
Donor ID: H21.33.009 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 65
Donor ID: H21.33.010 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 93
Donor ID: H21.33.011 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 83
Donor ID: H21.33.012 | APOE Genotype: 2/4 | CERAD score: Sparse | Cognitive Status: Dementia | Age at death: 93
Donor ID: H21.33.013 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 94
Donor ID: H21.33.014 | APOE Genotype: 2/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 92
Donor ID: H21.33.015 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 98
Donor ID: H21.33.016 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: Dementia | Age at death: 94
Donor ID: H21.33.017 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 92
Donor ID: H21.33.018 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 89
Donor ID: H21.33.019 | APOE Genotype: 2/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 75
Donor ID: H21.33.020 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 82
Donor ID: H21.33.021 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 99
Donor ID: H21.33.022 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 82
Donor ID: H21.33.023 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 102
Donor ID: H21.33.025 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 88
Donor ID: H21.33.026 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 90
Donor ID: H21.33.027 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 92
Donor ID: H21.33.028 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 72
Donor ID: H21.33.029 | APOE Genotype: 2/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 89
Donor ID: H21.33.030 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 89
Donor ID: H21.33.031 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 84
Donor ID: H21.33.032 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 98
Donor ID: H21.33.033 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 83
Donor ID: H21.33.034 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 90
Donor ID: H21.33.035 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 97
Donor ID: H21.33.036 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 93
Donor ID: H21.33.037 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 88
Donor ID: H21.33.038 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 84
Donor ID: H21.33.039 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 88
Donor ID: H21.33.040 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 83
Donor ID: H21.33.041 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 98
Donor ID: H21.33.042 | APOE Genotype: 4/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 91
Donor ID: H21.33.043 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: Dementia | Age at death: 95
Donor ID: H21.33.044 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 88
Donor ID: H21.33.045 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 94
Donor ID: H21.33.046 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 97
Donor ID: H21.33.047 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 90
```

In [8]:

```python
#3. Print sorted list of patients
cerad_order = {"Absent": 0, "Sparse": 1, "Moderate": 2, "Frequent": 3}
Patient.all_patients.sort(
    key=lambda p: cerad_order.get(p.CERAD_score.strip(), 99)  # 99 = "unknown" catch
)
# Print sorted list
print("\nPatients sorted by CERAD score:")
```

```
for patient in Patient.all_patients:
    print(patient, getattr(patient, "CERAD_score", "N/A"))
```

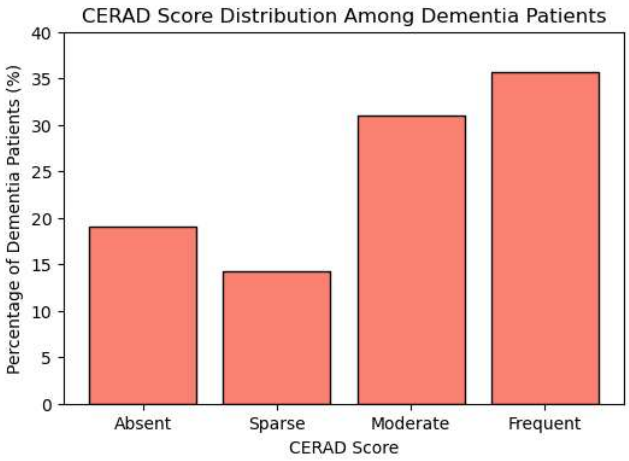
```
Patients sorted by CERAD score:
Donor ID: H19.33.004 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 80 Absent
Donor ID: H20.33.002 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 97 Absent
Donor ID: H20.33.012 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 91 Absent
Donor ID: H20.33.034 | APOE Genotype: 2/2 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 85 Absent
Donor ID: H20.33.035 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 99 Absent
Donor ID: H20.33.044 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 81 Absent
Donor ID: H21.33.001 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: Dementia | Age at death: 80 Absent
Donor ID: H21.33.003 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 78 Absent
Donor ID: H21.33.004 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 93 Absent
Donor ID: H21.33.011 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 83 Absent
Donor ID: H21.33.016 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: Dementia | Age at death: 94 Absent
Donor ID: H21.33.023 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 102 Absent
Donor ID: H21.33.028 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 72 Absent
Donor ID: H21.33.037 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 88 Absent
Donor ID: H21.33.038 | APOE Genotype: 3/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 84 Absent
Donor ID: H21.33.041 | APOE Genotype: 2/3 | CERAD score: Absent | Cognitive Status: No dementia | Age at death: 98 Absent
Donor ID: H20.33.001 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 82 Sparse
Donor ID: H20.33.014 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 82 Sparse
Donor ID: H20.33.019 | APOE Genotype: 3/4 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 87 Sparse
Donor ID: H20.33.024 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 90 Sparse
Donor ID: H20.33.043 | APOE Genotype: 4/4 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 85 Sparse
Donor ID: H21.33.006 | APOE Genotype: 3/4 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 97 Sparse
Donor ID: H21.33.012 | APOE Genotype: 2/4 | CERAD score: Sparse | Cognitive Status: Dementia | Age at death: 93 Sparse
Donor ID: H21.33.014 | APOE Genotype: 2/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 92 Sparse
Donor ID: H21.33.015 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 98 Sparse
Donor ID: H21.33.019 | APOE Genotype: 2/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 75 Sparse
Donor ID: H21.33.032 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: No dementia | Age at death: 98 Sparse
Donor ID: H21.33.043 | APOE Genotype: 3/3 | CERAD score: Sparse | Cognitive Status: Dementia | Age at death: 95 Sparse
Donor ID: H20.33.005 | APOE Genotype: 2/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 99 Moderate
Donor ID: H20.33.008 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 92 Moderate
Donor ID: H20.33.011 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 93 Moderate
Donor ID: H20.33.013 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 94 Moderate
Donor ID: H20.33.015 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 88 Moderate
Donor ID: H20.33.016 | APOE Genotype: 2/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 93 Moderate
Donor ID: H20.33.025 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 94 Moderate
Donor ID: H20.33.027 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 99 Moderate
Donor ID: H20.33.028 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 94 Moderate
Donor ID: H20.33.029 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 91 Moderate
Donor ID: H20.33.032 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 98 Moderate
Donor ID: H20.33.036 | APOE Genotype: 2/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 100 Moderate
Donor ID: H20.33.041 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 91 Moderate
Donor ID: H21.33.005 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 95 Moderate
Donor ID: H21.33.013 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 94 Moderate
Donor ID: H21.33.021 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 99 Moderate
Donor ID: H21.33.022 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 82 Moderate
Donor ID: H21.33.025 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 88 Moderate
Donor ID: H21.33.027 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 92 Moderate
Donor ID: H21.33.030 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 89 Moderate
Donor ID: H21.33.031 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 84 Moderate
Donor ID: H21.33.033 | APOE Genotype: 3/4 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 83 Moderate
Donor ID: H21.33.035 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: No dementia | Age at death: 97 Moderate
Donor ID: H21.33.039 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 88 Moderate
Donor ID: H21.33.042 | APOE Genotype: 4/4 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 91 Moderate
Donor ID: H21.33.046 | APOE Genotype: 3/3 | CERAD score: Moderate | Cognitive Status: Dementia | Age at death: 97 Moderate
Donor ID: H20.33.004 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 86 Frequent
Donor ID: H20.33.017 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 69 Frequent
Donor ID: H20.33.018 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 81 Frequent
Donor ID: H20.33.020 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 81 Frequent
Donor ID: H20.33.026 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 75 Frequent
Donor ID: H20.33.030 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 86 Frequent
Donor ID: H20.33.031 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 87 Frequent
Donor ID: H20.33.033 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 68 Frequent
Donor ID: H20.33.037 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 96 Frequent
Donor ID: H20.33.038 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 90 Frequent
Donor ID: H20.33.039 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 96 Frequent
Donor ID: H20.33.040 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 98 Frequent
Donor ID: H20.33.045 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 77 Frequent
Donor ID: H20.33.046 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 94 Frequent
Donor ID: H21.33.002 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 70 Frequent
Donor ID: H21.33.007 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 86 Frequent
Donor ID: H21.33.008 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 91 Frequent
Donor ID: H21.33.009 | APOE Genotype: 4/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 65 Frequent
Donor ID: H21.33.010 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 93 Frequent
Donor ID: H21.33.017 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 92 Frequent
Donor ID: H21.33.018 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 89 Frequent
Donor ID: H21.33.020 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 82 Frequent
Donor ID: H21.33.026 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 90 Frequent
Donor ID: H21.33.029 | APOE Genotype: 2/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 89 Frequent
Donor ID: H21.33.034 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 90 Frequent
Donor ID: H21.33.036 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 93 Frequent
Donor ID: H21.33.040 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 83 Frequent
Donor ID: H21.33.044 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 88 Frequent
Donor ID: H21.33.045 | APOE Genotype: 3/4 | CERAD score: Frequent | Cognitive Status: Dementia | Age at death: 94 Frequent
Donor ID: H21.33.047 | APOE Genotype: 3/3 | CERAD score: Frequent | Cognitive Status: No dementia | Age at death: 90 Frequent
```

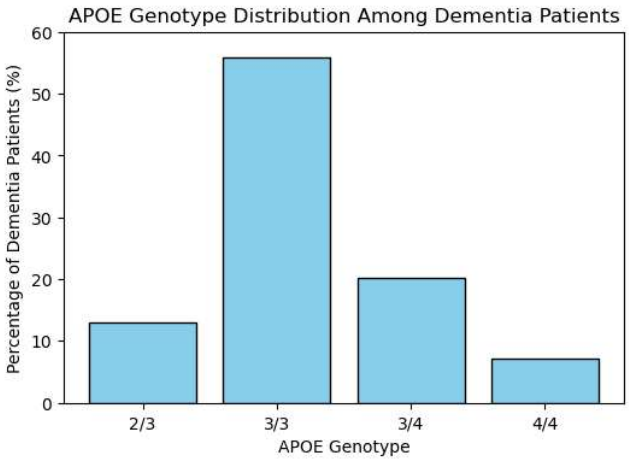
In [9]:
```
#4. Sorted dictionary
def sort_by_cerad(cls):
# Ensure dictionary is clean
    cls.cerad_groups.clear()
    for patient in cls.all_patients:
```

```
        cerad = patient.CERAD_score.strip()
        if cerad not in cls.cerad_groups:
            cls.cerad_groups[cerad] = []
        cls.cerad_groups[cerad].append(patient)

def subsort_by_age(cls):
    for cerad, patients in cls.cerad_groups.items():
        # Sort in place by numeric age (invalid -> big number at end)
        patients.sort(
            key=lambda p: int(p.age_at_death) if str(p.age_at_death).isdigit() else 999
        )
        cls.cerad_groups[cerad] = patients
```

In [10]:

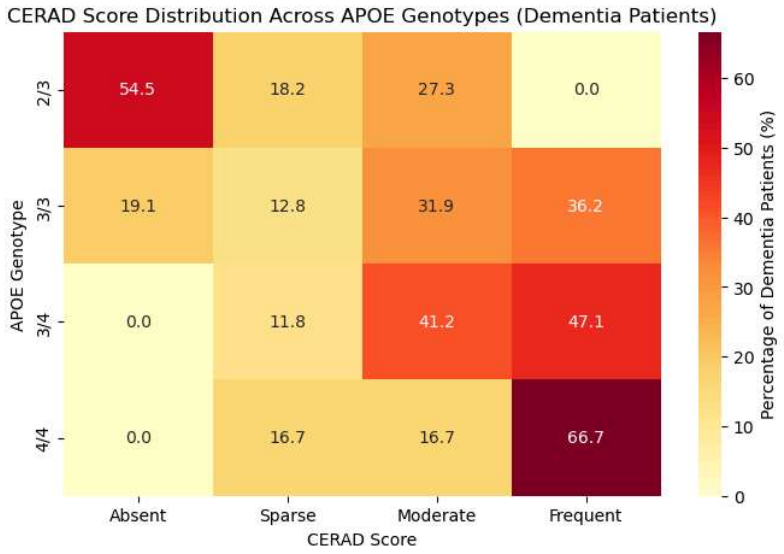
```
#5. Bar graph (CERAD scores vs number of patients)
# Filter only patients with dementia
dementia_patients = [
    p for p in Patient.all_patients
        if p.cognitive_status and "dementia" in p.cognitive_status.lower()
        and p.CERAD_score
    ]

    # Count how many dementia patients have each CERAD score
cerad_order = ["Absent", "Sparse", "Moderate", "Frequent"]
counts = {cerad: 0 for cerad in cerad_order}

for p in dementia_patients:
    cerad = p.CERAD_score.strip().title()
    if cerad in counts:
        counts[cerad] += 1

# Convert counts to percentages
total_dementia = len(dementia_patients)
percentages = {cerad: (count / total_dementia) * 100 for cerad, count in counts.items()}

# Plot
plt.figure(figsize=(6, 4))
plt.bar(percentages.keys(), percentages.values(), color="salmon", edgecolor="black")
plt.ylabel("Percentage of Dementia Patients (%)")
plt.xlabel("CERAD Score")
plt.title("CERAD Score Distribution Among Dementia Patients")
plt.ylim(0, 40)
plt.show()

#6. percentage of demntia patients vs apoe genotype
# Filter only patients with dementia
dementia_patients = [
    p for p in Patient.all_patients
    if p.cognitive_status and "dementia" in p.cognitive_status.lower()
    and p.apoe
]

# Count how many dementia patients have each APOE genotype
apoe_order = ["2/3", "3/3", "3/4", "4/4"]
counts = {apoe: 0 for apoe in apoe_order}

for p in dementia_patients:
    genotype = p.apoe.strip()
    if genotype in counts:
        counts[genotype] += 1

# Convert counts to percentages
total_dementia = len(dementia_patients)
percentages = {genotype: (count / total_dementia) * 100 for genotype, count in counts.items()}

# Plot
plt.figure(figsize=(6, 4))
plt.bar(percentages.keys(), percentages.values(), color="skyblue", edgecolor="black")
plt.ylabel("Percentage of Dementia Patients (%)")
plt.xlabel("APOE Genotype")
plt.title("APOE Genotype Distribution Among Dementia Patients")
plt.ylim(0, 60)
plt.show()
```





In [11]:
```python
#7. APOE Genotype vs CERAD scores in the set of dementia patients
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Filter only patients with dementia
dementia_patients = [
    p for p in Patient.all_patients
    if p.cognitive_status and "dementia" in p.cognitive_status.lower()
    and p.apoe and p.CERAD_score
]

# Create DataFrame
data = []
for p in dementia_patients:
    data.append({
        "APOE": p.apoe.strip(),
        "CERAD": p.CERAD_score.strip().title()
    })

df = pd.DataFrame(data)

# Crosstab: rows = APOE, columns = CERAD, values = counts
crosstab = pd.crosstab(df["APOE"], df["CERAD"])

# Ensure all CERAD and APOE categories are included
cerad_order = ["Absent", "Sparse", "Moderate", "Frequent"]
apoe_order = ["2/3", "3/3", "3/4", "4/4"]

for cerad in cerad_order:
    if cerad not in crosstab.columns:
        crosstab[cerad] = 0

crosstab = crosstab[cerad_order]  # reorder columns
crosstab = crosstab.reindex(apoe_order)  # reorder rows

# Convert counts to percentages per APOE genotype
percent_crosstab = crosstab.div(crosstab.sum(axis=1), axis=0) * 100

# Plot heatmap
plt.figure(figsize=(8, 5))
sns.heatmap(
    percent_crosstab,
    annot=True, fmt=".1f", cmap="YlOrRd",
    cbar_kws={'label': 'Percentage of Dementia Patients (%)'}
)
```

```
plt.xlabel("CERAD Score")
plt.ylabel("APOE Genotype")
plt.title("CERAD Score Distribution Across APOE Genotypes (Dementia Patients)")
plt.show()
```



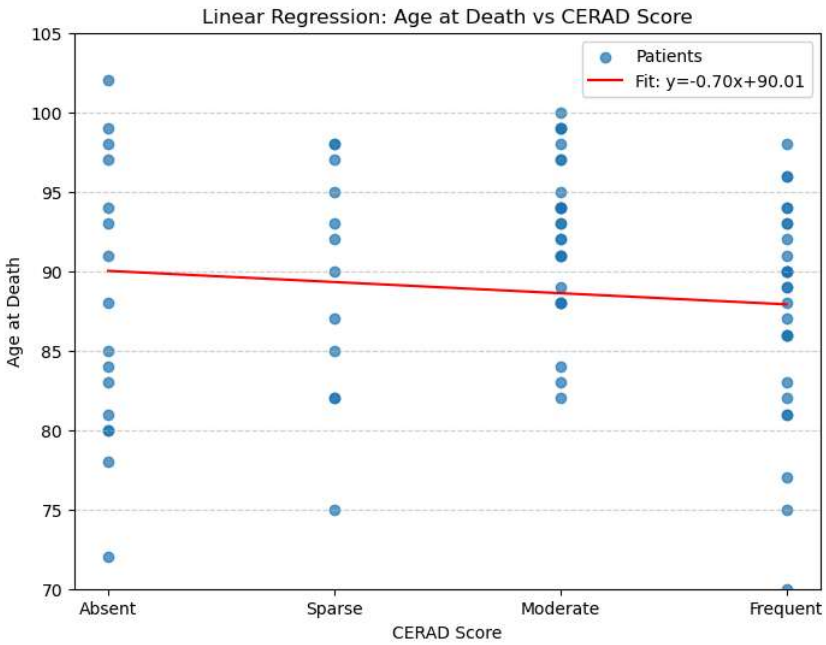
In [12]:
```
#8. Making a two-way ANOVA table
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols

data = pd.DataFrame([{
    "DonorID": p.donor_id,
    "APOE": p.apoe,
    "CERAD": p.CERAD_score,
    "AgeAtDeath": float(p.age_at_death) if p.age_at_death else None
    }
for p in Patient.all_patients])

#Drop rows with missing values (important for ANOVA)
data = data.dropna(subset=["APOE", "CERAD", "AgeAtDeath"])

#Make sure categorical variables are categorical
data["APOE"] = data["APOE"].astype("category")
data["CERAD"] = data["CERAD"].astype("category")

#Fit two-way ANOVA model
model = ols('AgeAtDeath ~ C(APOE) * C(CERAD)', data=data).fit()
anova_table = sm.stats.anova_lm(model, typ=2)

print("\nTwo-Way ANOVA Results:")
print(anova_table)

#Creating Linear Regression: Age at Death vs CERAD scores
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats

cerad_numeric_map = {"Absent": 0, "Sparse": 1, "Moderate": 2, "Frequent": 3}

x_cerad, y_cerad = [], []
for p in Patient.all_patients:
    try:
        if p.CERAD_score in cerad_numeric_map and p.age_at_death:
            x_cerad.append(cerad_numeric_map[p.CERAD_score])
            y_cerad.append(float(p.age_at_death))
    except (ValueError, TypeError):
        continue
```

```
Two-Way ANOVA Results:
                     sum_sq    df             F    PR(>F)
C(APOE)          5.472574e+01   5.0  1.988912e-01  0.938115
C(CERAD)        -1.554344e-11   3.0 -9.414988e-14  1.000000
C(APOE):C(CERAD)  1.491327e+03  15.0  1.806656e+00  0.059554
Residual          3.742097e+03  68.0           NaN       NaN
```

```
c:\Users\akrem\anaconda3\Lib\site-packages\statsmodels\base\model.py:1894: ValueWarning: covariance of constraints does not have full rank. The number of constraints is 5, but rank is 4
  warnings.warn('covariance of constraints does not have full '
c:\Users\akrem\anaconda3\Lib\site-packages\statsmodels\base\model.py:1894: ValueWarning: covariance of constraints does not have full rank. The number of constraints is 15, but rank is 13
  warnings.warn('covariance of constraints does not have full '
```

In [13]:
```
#9. Creating Linear Regression: Age at Death vs CERAD scores
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
```

```
cerad_numeric_map = {"Absent": 0, "Sparse": 1, "Moderate": 2, "Frequent": 3}

x_cerad, y_cerad = [], []
for p in Patient.all_patients:
    try:
        if p.CERAD_score in cerad_numeric_map and p.age_at_death:
            x_cerad.append(cerad_numeric_map[p.CERAD_score])
            y_cerad.append(float(p.age_at_death))
    except (ValueError, TypeError):
        continue

#Fit regression
slope, intercept, r_value, p_value, std_err = stats.linregress(x_cerad, y_cerad)

#Regression line
x_vals = np.array(sorted(set(x_cerad)))
y_vals = intercept + slope * x_vals

plt.figure(figsize=(8, 6))
plt.scatter(x_cerad, y_cerad, alpha=0.7, label="Patients")
plt.plot(x_vals, y_vals, color="red", label=f"Fit: y={slope:.2f}x+{intercept:.2f}")
plt.xticks(range(4), ["Absent", "Sparse", "Moderate", "Frequent"])
plt.xlabel("CERAD Score")
plt.ylabel("Age at Death")
plt.title("Linear Regression: Age at Death vs CERAD Score")
plt.ylim(70, 105)
plt.grid(axis="y", linestyle="--", alpha=0.6)
plt.legend()
plt.show()

print(f"Slope: {slope:.3f}")
print(f"Intercept: {intercept:.3f}")
print(f"R-squared: {r_value**2:.3f}")
print(f"P-value: {p_value:.3g}")
```




```
Slope: -0.703
Intercept: 90.014
R-squared: 0.010
P-value: 0.376
```

In [14]:
```
#10. EXPORT DEMENTIA PATIENT DATA TO A .CSV FILE
import pandas as pd

# Filter dementia patients
dementia_patients = [
    p for p in Patient.all_patients
    if p.cognitive_status and "dementia" in p.cognitive_status.lower()
    and p.apoe and p.CERAD_score
]

# Prepare lists for DataFrame
apoe_list = [p.apoe.strip() for p in dementia_patients]
cerad_list = [p.CERAD_score.strip().title() for p in dementia_patients]

# Optional: you can add numeric mapping for CERAD if needed
cerad_mapping = {"Absent": 0, "Sparse": 1, "Moderate": 2, "Frequent": 3}
cerad_numeric = [cerad_mapping.get(c, None) for c in cerad_list]

# Create DataFrame
df = pd.DataFrame({
    'APOE Genotype': apoe_list,
```

```
    'CERAD Score': cerad_list,
    'CERAD Numeric': cerad_numeric
})

# Write to CSV
df.to_csv('dementia_patients.csv', index=False)

print("CSV file 'dementia_patients.csv' has been created.")
```

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

Verify and validate your analysis:

To verify that our analyses returned believeable answers, we looked at the results of our statistical tests (two-way ANOVA) and figures to make sure there were no coding errors that made the figures incacurate/exhibit obvious problems. Referencing literature further helped us validate our results. For example, our two-way ANOVA returned a p-value of >0.05, meaning that there is no statistically significant effect of APOE genotypes or CERAD score on age of death. Notably, however, the p-value was close to 0.05 (0.059554) when assesing effect of APOE genotype on age of death across different CERAD scores. Various articles (in references section below) suggest that APOE 4 genotype is a high risk factor for Alzheimer's disease. When it comes to mortality rates, one 2023 Science Reports paper found that APOE 4 was associated with high mortality when someone had high amounts of "Alzheimer's-type neuropathology" but low mortality when they had low amounts of Alzheimer's-type neuropathology. Many studies also point to the fact that it lowers age of onset but don't specify conclusions about age of death. This goes against what we found from our tests. From papers from The Lancet Healthy Longevity and The Journal of Neuropathology and Experimental Neurology (cited below), CERAD scores are good predictors of Alzheimer's pathogenesis, but information about mortality is not specified. Thus, we can't use the literature to definitvely support our results from CERAD scores.

Conclusions and Ethical Implications:

The results we got (that APOE genotype and CERAD scores don't have an effect on age of death) indicates that the factors we explored don't correlate to a patient's health. However, the p-value from the ANOVA test that was close to 0.05 may point us to the fact that a longitudinal study with a bigger dataset may be better at getting similar results as seen in the literature. In terms of ethical implications, what we found states that although certain patients may have a certain genotype or test score, we can't conclusively say that their experience with Alzheimer's disease will be different than patients with different demographics.

Limitations and Future Work:

As mentioned earlier, our conclusion is that it is inconclusive whether or not APOE genotype and CERAD score correlate to different patient health and mortality outcomes. The size of the dataset as well as the fact that it came from a non-longitudinal study may be effecting the significance level found from our tests, so something to pursue in the future is testing the same parameters on a longitudinal study with a large dataset. In the future, it may be beneficial to further test the correlation between different factors and mortality from Alzheimer's because it could shape new ideas and directions for treatment developments.

Questions for the TA:

No questions at this time.

References

Alberto Serrano-Pozo, Jing Qian, Alona Muzikansky, Sarah E Monsell, Thomas J Montine, Matthew P Frosch, Rebecca A Betensky, Bradley T Hyman, Thal Amyloid Stages Do Not Significantly Impact the Correlation Between Neuropathological Change and Cognition in the Alzheimer Disease Continuum, Journal of Neuropathology & Experimental Neurology, Volume 75, Issue 6, June 2016, Pages 516–526, https://doi.org/10.1093/jnen/nlw026

Emma Nichols, Richard Merrick, Simon I Hay, Dibya Himali, Jayandra J Himali, Sally Hunter, Hannah A D Keage, Caitlin S Latimer, Matthew R Scott, Jaimie D Steinmetz, Jamie M Walker, Stephen B Wharton, Crystal D Wiedner, Paul K Crane, C Dirk Keene, Lenore J Launer, Fiona E Matthews, Julie Schneider, Sudha Seshadri, Lon White, Carol Brayne, Theo Vos, The prevalence, correlation, and co-occurrence of neuropathology in old age: harmonisation of 12 measures across six community-based autopsy studies of dementia, The Lancet Healthy Longevity, Volume 4, Issue 3, 2023, Pages e115-e125, ISSN 2666-7568, https://doi.org/10.1016/S2666-7568(23)00019-3.

OpenAI. (2025). ChatGPT (Sept 30 version) [Large language model]. https://chat.openai.com/

In our project, ChatGPT was used to help us write the code to generate our figures (bar graphs, heat map, linear regression) and our statistical analyses/tests (t-test, ANOVA). ChatGPT was also used to debug our code. Examples of prompts include "how to code two-way ANOVA in Python" or pasting our code and asking it how to complete certain coding tasks/what issues there were.

Pirraglia, E., Glodzik, L. & Shao, Y. Lower mortality risk in APOE4 carriers with normal cognitive ageing. Sci Rep 13, 15089 (2023). https://doi.org/10.1038/s41598-023-41078-5

Sando, S.B., Melquist, S., Cannon, A. et al. APOE ε4 lowers age at onset and is a high risk factor for Alzheimer's disease; A case control study from central Norway. BMC Neurol 8, 9 (2008). https://doi.org/10.1186/1471-2377-8-9