

Module_1: Alzheimer's Disease

Team Members:

Daniel Garcia-Soliz and Meredith Lineweaver

Project Title:

Alzheimer's Disease and Gender

Project Goal:

This project seeks to analyze the correlation between other diseases and Alzheimer's impacting neurological impairment by considering MMSE score.

Disease Background:

Fill in information about 11 bullets:

- Prevalence & incidence:
 - Over 7 million Americans have Alzheimer's Disease.
 - 1 in 9 people that are 65 years of age or older.
 - 200,000 American have younger-onset dementia (age 30-64 years)
 - Approximately 500,000 people were diagnosed with Alzheimer's in 2024
- Economic burden:
 - Global cost by 2030 is expected to be 2 trillion for care. Dependent on the severity of the disease, the annual range for one person in America is a wide. For mild AD, it is 468.28 and for severe AD it is 138,023.97 per person per a year. These are based off the total (direct and indirect) treatment and "loss of treatment time".
(<https://www.sciencedirect.com/science/article/pii/S2212109923000948?via%3Dihub#sec3>)
- Risk factors (genetic, lifestyle)
 - 2/3 of Americans with Alzheimer's are women.
 - Older Black women are almost 2x as likely and older Hispanics are 1.5x as likely to have Alzheimer's or other dementias as older whites.
 - The greatest risk if increasing age.
 - Family history, the risk increase with each family member that has the illness.
 - Bad sleep habits, smoking, hypertension, and diabetes.

- Head injury.
- Societal determinants:
 - Lower education level
 - Chronic stress
 - Social isolation
 - Poor economic stability
 - Inconsistent or inadequate access to healthcare
- Symptoms:
 - Memory loss that disrupts daily life
 - Difficulty planning or solving problems
 - Trouble completing basic/familiar task
 - Confusion
 - New problems with words in speaking or writing
 - Changes in mood and personality
 - Misplacing things
 - Decreased or poor judgement with decisions
 - Social withdrawal
- Diagnosis:
 - Begins with acknowledgement of symptoms, current lifestyle patterns, medications, history, overall health, and changes in behavior. With enough suspicion or request from the patient it is common to carry through with testing. Memory/Problem solving testing. Fluid testing to rule out other possible factors. Measure proteins associated with Alzheimers in CSF (cerebrospinal fluid). CT, PET, and MRI to solidify diagnosis from other tests. Looking for degradation of brain. (<https://www.nia.nih.gov/health/alzheimers-symptoms-and-diagnosis/how-alzheimers-disease-diagnosed>)
- Standard of care treatments (& reimbursement):
 - 12 million Americans provide unpaid care for a family member or friend with dementias.
 - The value of unpaid care hours is valued at \$413.5 billion.
 - 30% of caregivers are age 65 or older.
 - Over 1/3 of dementia caregivers are daughters.
 - Most caregivers live with the person with dementia.
 - 70% of the lifetime cost of caring for someone with dementia is covered by families.
 - There are not enough dementia care specialists in 55% of communities.
- Disease progression & prognosis:
 - 1/3 older Americans dies with Alzheimer's or another dementia.
 - 6th leading cause of death among people age 65 and older in 2022.
 - 5 stages: preclinical, mild cognitive impairment, mild dementia, moderate dementia, severe dementia
 - Alzheimer's disease is often diagnosed in the mild dementia stage

- People with Alzheimer's live between 3 and 11 years after diagnosis, but some can live 20 years or more.
- Prognosis depends on how far the disease has advanced when diagnosed.
- Untreated vascular risk factors, like high blood pressure, are associated with a faster rate of progression of the disease.
- Continuum of care providers:
 - Beginning with a primary care provider suspicion following questioning and gathering of previous medical history. After the diagnosis it is useful to set up with both neurologists near by and other memory supportive teams to assist with individual living patterns. The most effective continuum of care providers are communities or organizations which support the patients with constant contact and the ability to go to their home.
(<https://www.hospicechesapeake.org/2022/06/continuum-of-care-offers-dementia-patients-better-quality-of-life/>)
- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology)
 - Many correlations with biological mechanisms typically come from biological markers within the brain. Understanding some of the basic cellular pathophysiology allows a better understanding on diagnosis the severity for AD or other types of dementia. Based on Dogan and Kocahan, low-level activation of NMDA, amyloid-beta, BDNF, etc
(<https://pmc.ncbi.nlm.nih.gov/articles/PMC5290713/>).
- Clinical Trials/next-gen therapies
 - There varies different examinations for alzheimers. Impacts of certain gene expressions. Neurological examinations by sociological questioning. For next-gen therapies, any anti-amyloid or anti-body drugs for inhibition of continuing growths of clots/plaque.

Sources: <https://www.alz.org/alzheimers-dementia/facts-figures>
<https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-stages/art-20048448>

Data-Set:

The source of this data is from a paper published on October 14, 2024 by nature neuroscience titled "Integrated multimodal cell atlas of Alzheimer's disease". There are two data sets for this project. Both data sets studied 84 donors, 51 female and 33 male. The data set titled UpdatedMetaData.csv provides demographic and medical history of the donors. The file named UpdatedLuminex.csv provides a ratio for ABeta40, ABeta42, tTAU, and pTAU mass. The authors of this study found their results by generating quantitative neuropathological measurement, single-nucleus RNA sequencing, and cellularly resolved spatial transcriptomics. The profiled 3.4 million high-quality nuclei,

mapping each one to a molecular cell type from the BRAIN Initiative MTG cellular taxonomy.

Link to the source of the data sets: <https://doi.org/10.1038/s41593-024-01774-5>

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

class Patient:
    all_patients = []

    def __init__(self, DonorID, sex=None, death_age=None, cog_stat=None,
                  consensus_dx=None, brain_weight=None, mmse_score=None):
        self.DonorID = DonorID
        self.sex = sex
        self.death_age = death_age
        self.cog_stat = cog_stat
        self.consensus_dx = consensus_dx # list of diagnoses or None
        self.brain_weight = brain_weight
        self.mmse_score = mmse_score # last MMSE score (integer)
        Patient.all_patients.append(self)

    def __repr__(self):
        dx_display = self.consensus_dx if self.consensus_dx else "None"
        return (f"{self.DonorID} | sex: {self.sex} | Death Age {self.death_age} | "
                f"Cognitive Status {self.cog_stat} | Consensus Dx {dx_display} | "
                f"Brain Wt {self.brain_weight} | Last MMSE: {self.mmse_score}")

    @classmethod
    def instantiate_from_csv(cls, filename: str):
        with open(filename, encoding="utf8") as f:
            reader = csv.DictReader(f)
            headers = reader.fieldnames

            for row in reader:
                donor_id = row["Donor ID"].strip()
                sex = row["Sex"].strip() if row["Sex"] else None
                death_age = int(row["Age at Death"]) if row["Age at Death"] else None
                cog_stat = row["Cognitive Status"].strip() if row["Cognitive Status"] else None

                # collect diagnoses marked "Checked"
                consensus_cols = [h for h in headers if h.startswith("Consensus Clinical Dx (choice)")]
                diagnoses = []
                for col in consensus_cols:
                    val = row[col].strip()
                    if val == "Checked":
                        dx_name = col.replace("Consensus Clinical Dx (choice)", "")
                        diagnoses.append(dx_name)
                if not diagnoses:
                    diagnoses = None
```

```

        # brain weight
        brain_weight = None
        if row["Fresh Brain Weight"]:
            try:
                brain_weight = float(row["Fresh Brain Weight"])
            except ValueError:
                brain_weight = None

    # last MMSE score
    mmse_score = None
    if "Last MMSE Score" in row and row["Last MMSE Score"]:
        try:
            mmse_score = int(row["Last MMSE Score"])
        except ValueError:
            mmse_score = None

    # create patient object
    Patient(
        DonorID=donor_id,
        sex=sex,
        death_age=death_age,
        cog_stat=cog_stat,
        consensus_dx=diagnoses,
        brain_weight=brain_weight,
        mmse_score=mmse_score
    )

# -----
# Load Patients
# -----
Patient.instantiate_from_csv("UpdatedMetaData.csv")
Patient.all_patients.sort(key=lambda p: p.consensus_dx or [], reverse=False)

# -----
# Count No dementia vs Dementia
# -----
print("\nTotal Patients with Dementia:")
no_dementia_count = sum(1 for p in Patient.all_patients if p.cog_stat == "No de
dementia_count = sum(1 for p in Patient.all_patients if p.cog_stat != "No de
total_patients = len(Patient.all_patients)

print(f"\nNumber of 'No dementia' patients: {no_dementia_count}")
print(f"Number of 'Dementia/Other' patients: {dementia_count}")
print(f"Total patients: {total_patients}")

# -----
# Count each individual consensus diagnosis
# -----
from collections import Counter

dx_counter = Counter()

for p in Patient.all_patients:
    if p.consensus_dx: # skip if None

```

```

dx_counter.update(p.consensus_dx)

print("\nConsensus diagnosis counts:")
for dx, count in dx_counter.items():
    print(f"{dx}: {count}")

# -----
# Bar chart: Alzheimer's only vs Alzheimer's+Other vs Control
# -----
group_counts = {"Alzheimer's only": 0, "Alzheimer's + Other": 0, "Control": 0}

for p in Patient.all_patients:
    if not p.consensus_dx:
        continue

    dx_list = p.consensus_dx

    # Control
    if "Control" in dx_list:
        group_counts["Control"] += 1

    # Alzheimer's categories
    elif any(dx in ["Alzheimers disease", "Alzheimers Possible/ Probable"] for dx in dx_list):
        if len(dx_list) == 1: # only AD
            group_counts["Alzheimer's only"] += 1
        else: # AD plus at least one other
            group_counts["Alzheimer's + Other"] += 1

# Plotting
plt.figure(figsize=(8, 6))
plt.bar(group_counts.keys(), group_counts.values(),
        color=["skyblue", "lightcoral", "lightgreen"])
plt.ylabel("Number of Patients")
plt.title("Patient Counts: Alzheimer's Only vs Alzheimer's+Other vs Control")

# add counts on bars
for i, v in enumerate(group_counts.values()):
    plt.text(i, v + 0.5, str(v), ha='center', fontweight='bold')

plt.grid(axis="y", linestyle="--", alpha=0.7)
plt.tight_layout()
plt.show()

# -----
# ANOVA: Compare Last MMSE across 3 groups
# -----
alz_only_scores = []
alz_plus_other_scores = []
control_scores = []

for p in Patient.all_patients:
    if p.mmse_score is None or not p.consensus_dx:
        continue
    dx_list = p.consensus_dx

```

```

# Control
if "Control" in dx_list:
    control_scores.append(p.mmse_score)

# Alzheimer's categories
elif any(dx in ["Alzheimers disease", "Alzheimers Possible/ Probable"]) and len(dx_list) > 1:
    if len(dx_list) == 1: # only AD
        alz_only_scores.append(p.mmse_score)
    else: # AD + something else
        alz_plus_other_scores.append(p.mmse_score)

print("\nGroup sizes and means (Last MMSE):")
print(f"Alzheimer's only: n={len(alz_only_scores)}, mean={np.mean(alz_only_scores)}")
print(f"Alzheimer's + Other: n={len(alz_plus_other_scores)}, mean={np.mean(alz_plus_other_scores)}")
print(f"Control: n={len(control_scores)}, mean={np.mean(control_scores)}")

if alz_only_scores and alz_plus_other_scores and control_scores:
    f_stat, p_value = stats.f_oneway(alz_only_scores, alz_plus_other_scores, control_scores)
    print("ANOVA F-statistic:", f_stat)
    print("ANOVA p-value:", p_value)

    plt.figure(figsize=(8, 6))
    plt.boxplot([alz_only_scores, alz_plus_other_scores, control_scores],
                tick_labels=["Alzheimer's only", "Alzheimer's + Other", "Control"],
                patch_artist=True,
                boxprops=dict(facecolor='skyblue', color='blue'),
                medianprops=dict(color='red'))
    plt.title("Last MMSE Score Across Groups\nANOVA p-value = {:.4f}".format(p_value))
    plt.ylabel("Last MMSE Score")
    plt.grid(True, axis='y', linestyle='--', alpha=0.7)
    plt.tight_layout()
    plt.show()
else:
    print("⚠ Not enough data in one or more groups to run ANOVA.")

# 9 GET PATIENT ATTRIBUTES THAT WE WANT TO COMPARE ON A SCATTER PLOT

death_age_list = []
mmse_scores = []

for patient in Patient.all_patients:
    if patient.death_age is not None and patient.mmse_score is not None:
        death_age_list.append(patient.death_age)
        mmse_scores.append(patient.mmse_score)

X = death_age_list # Independent variable
y = mmse_scores    # Dependent variable

print("Ages at death:", X)
print("MMSE scores:", y)

# 10 VISUALIZE DATA ON A SCATTER PLOT

plt.figure(figsize=(8, 6))
plt.scatter(X, y, alpha=0.7)
plt.xlabel('Age at Death')
plt.ylabel('Last MMSE Score')

```

```

plt.title('Scatter Plot of Age at Death vs Last MMSE Score')
plt.grid(True, linestyle='--', alpha=0.6)
plt.show()

# 11 EXPORT DATA TO A .csv FILE

import pandas as pd

df = pd.DataFrame({
    'Age at Death': X,
    'Last MMSE Score': y
})

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

# -----
# Linear Regression: Age of Death vs Last MMSE
# -----
from sklearn.linear_model import LinearRegression

# Clean data (remove None values)
clean_data = [(p.death_age, p.mmse_score) for p in Patient.all_patients
               if p.death_age is not None and p.mmse_score is not None]

if clean_data:
    X, y = zip(*clean_data)
    X = np.array(X).reshape(-1, 1)
    y = np.array(y)

    # Fit model
    model = LinearRegression()
    model.fit(X, y)

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

    # Plot scatter
    plt.figure(figsize=(8, 6))
    plt.scatter(X, y, label="Data", alpha=0.7)

    # Plot regression line
    x_range = np.linspace(min(X), max(X), 100).reshape(-1, 1)
    plt.plot(x_range, model.predict(x_range), color="red", label="Regression Line")

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

    plt.xlabel("Age of Death")
    plt.ylabel("Last MMSE Score")
    plt.title("Age of Death vs Last MMSE Score (with Regression Line)")
    plt.legend()
    plt.grid(True, linestyle="--", alpha=0.7)
    plt.tight_layout()

```



```
plt.show()  
else:  
    print("⚠️ No valid data for regression (missing Age or MMSE).")
```

Verify and validate your analysis:

To verify our hypothesis, we created a bar graph to compare Alzheimer's patients that have other conditions with patients that do not. We then compared the MMSE scores of those two groups. We conducted an ANOVA test to determine the significance between these groups. These results are consistent with our initial predictions. While having certain conditions can increase the lifetime risk of developing Alzheimer's Disease, there is currently no link to those conditions making Alzheimer's symptoms worse. This was evident in our analysis that there was no correlation between MMSE score and other conditions (<https://my.clevelandclinic.org/health/diseases/9164-alzheimers-disease>). Our linear regression model suggests that there is little to no correlation between MMSE score and age of death. This finding is also consistent with existing literature as outlined in this article: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6211473/>

Conclusions and Ethical Implications:

(Think about the answer your analysis generated, draw conclusions related to your overarching question, and discuss the ethical implications of your conclusions.) The purpose of this notebook is to analyze whether different neurological disorders impacts the MMSE score. Based on Arevalo-Rodriguez et al., this stands for Mini-Mental State Examination: detection of cognitive impairment. (<https://pmc.ncbi.nlm.nih.gov/articles/PMC6464748/>) It was determined to not be a strong test by itself, but within an entirety of examinations may narrow down the severity of the neurological impacts. With a mix of sociological testing and biological markers, patients with probable Alzheimers can be more easily detected. The only contraindication toward diagnosis Alzheimers with any of these examinations is the confusion of other clinical diagnosis hidden in the patient. Comparing the MMSE score between patients with alzheimers and other diseases may provide a clear view on the similarites the broad dementia has with other neurological impairments.

This data set comes from 84 patient organ donations, analyzed with a plethora of tests prior to death. Important information includes: age, diagnosis age, death age, fresh brain weight, other diagnosis, biomarkers, other sociological examinations, etc. As seen in the count,the split between "Dementia/other" and "No dementia" are equally split with 42. Creating a larger funnel, there are 17 patients with diseases outside of Alzheimers, allowing further analysis of this impact on neurological defecit. The second graph demonstrates the range of MMSE scores, along with the p-value based off of an ANOVA score. It seems the null hypothesis is accepted (there is not a significant difference with patients of only alzheimers and other diseases when comparing MMSE score). The

control subjects are a comparison with patients of no diagnosed Alzheimers or other diseases. (ADD DATA REGARDING MEREDITHS ANALYSIS)

Based on the data and its comparison to the project goal, it seems to follow a pattern of no significant additions to neurological disorder when having other diseases with Alzheimers. When comparing these results with Bowler et al. it may help explain the reason (<https://pubmed.ncbi.nlm.nih.gov/9193204/>). While other diagnosis, such as vascular dementia, evolve in similar pattern, it may not have an additive effect. Instead, the disorder may be apart of the Alzheimers or easily confused in with the other examinations. Furthermore, MMSE score seeks out mild dementia like disorder, meaning that there being no significant difference within the three groups may support the use of the test.

Limitations and Future Work:

(Think about the answer your analysis generated, draw conclusions related to your overarching question, and discuss the ethical implications of your conclusions.

Based on our original question we saw that MMSE score does not have a significant correlation between AD and other diseases. Since MMSE score cannot be removed from a typical diagnosis, it would be helpful to use this as a support of other examinations for AD or other dementia based disorders. Clinicians should focus on not basing AD on sociological markers but instead a mix with biomarkers such as beta-amyloid. This may create difficulties with patients on analysis. Using this analysis, many online websites regarding sociological examinations such as this must have caution with the description of the scoring. It may be worrying as a patient to have slightly less MMSE score than the average human when in reality the difference does not define potential AD.

NOTES FROM YOUR TEAM:

This is where our team is taking notes and recording activity.

QUESTIONS FOR YOUR TA:

These are questions we have for our TA.

In []: `#Our code is having some problems running and creating both bars on the graph
#Also we think our question will not have enough data associated with our qu`