In Wave 1，

**Demographical Relationship findings**:

**\*\*Gender\*\***: females have better recall index than male

**\*\*Race\*\***: 3 levels; white caucasians have the the best recall index,black have the worst, other races are in between.

**\*\*Census Region\*\***: among all well-known regions, a little difference in recall index is observed between regions

**\*\*Education Years\*\***: the recall index improves with educational years; the increment diminishes as the years increases; high school graduates (12~13 years) roughly have the same recall performance as bachelors (13-17 years)

**\*\*Current Marital Status\*\***: very little difference in between different maritial status

**\*\*Age\*\***: recall index slowly decreases with the age, from a median of 16 words for age in 23~49 to 61.85 to 11 words in age 61 to 85

**\*\*Living Siblings\*\***: non-linear relationships between the recall and count of living siblings, the best performance occurs at 0 and 12 siblings, and worst performance occur at 6 silings

**\*\*Children Born\*\***: slowly declining performances with the number of children ever borned

Economic Variables Relationship findings:

1. Individual earnings (RwEARN) is positively correlated with recall index
2. Total Household income (RwITOT) is positively correlated with recall index
3. Net value of vehicles (RwATRAN) is positively correlated with recall index
4. Value of chequng account (RwACHCK) is positively correlated with recall index
5. Value of primary residence (RwAHOUS) is positively correlated with recall index

Occupational variables Relationship Findings:

1. If this job is physically demanding (stooping, lifting, kneeling), cognitive performance declines as the intensity increase.
2. If this job is cognitively demanding (stress, eyesight), cognitive performance improves as the intensity increase.
3. Systematic difference exist between occupation categories.

Functional Limitations:

1. RwWALK: non-linear relationship with recall index
2. RwBATH: non-linear relationship with recall index
3. RwBED: non-linear relationship with recall index
4. RwTOILTH: non-linear relationship with recall index
5. RwPHONE: non-linear relationship with recall index
6. RwMAP: non-linear relationship with recall index
7. RwSHOP: non-linear relationship with recall index

**Implication to Modelling**

Demographics:

1. Age: center it around age 65 years, treat as a continuous variable

2. Years of education: fit a non-linear transformation, consider log(years) and sqrt(years)

3. Number of Living Siblings: discretize into none (0), few (1-3), several (4-8), many (>9)

4. Number of Children Ever Born: discretize into none (0), few (1-3), several (4-8), many (>9)

Occupation Variables:

1. Discretize physical and psychological requirement variables into high (1) and low (0)
2. Further categorize occupations into physical, intellectual, and social works

Economic Variable

1. Log transform all correlated economic variables

Functional Limitations:

RwWALK, RwBATH, RwBED, RwTOILTH, RwPHONE, RwMAP, RwSHOP

Questions for the mentor:

1. **How to connect back to big data:**
   1. connection to **broader** aging-related implications
      1. Just explain how our research is unique and valuable to contribute to the health research
      2. Justify our methods (i.e LSTM)
      3. Change age baseline to 65, or other ages of disease onset
   2. The gaps when explicitly **justifying** the statistical methods and AI modeling
      1. Sensor data are under regulations
      2. Data integration issues: how should tech be implemented
2. Poster presentation format? Suggestions on deliverables?
   1. Poster and on-stage presentations
      1. Poster: basic idea of what your research
      2. On-stage: a small pitch competition; 2-3 minutes; 2-3 people maximum, 1 person preferred
   2. 20 groups
   3. Ranking by judges， judges are consultants, data scientists
   4. Manuscript, more than 20 pages
      1. Scope
      2. Landscape: weakness, strengths
      3. **Strategic goals**
      4. Explain EDA, ML models,
      5. How it contributes to health research, **unique innovations** and justifications
3. Is it worth researching to extrapolate the cognitive decline over time for subject of different conditions? What should be the deliverable?  
   Detect threshold in work hours
   1. Forecast progression
   2. MCAC matrix
   3. 5-year decline score
   4. Risk table
   5. Divide people into subgroups—Divide into low-educated jobs (physically demanding), highly-educated jobs, and display subgroup tree
      1. Industry-heavy jobs, soldiers, fire fighters
4. How can we connect the question to health care system efficiency? How about social cost?
5. Recommendation of AI tools and usages
   1. Guru model—LSTM model input gate and output gate
   2. Overleaf: Zotero integration

<https://www.news-medical.net/news/20250606/New-retinal-prosthesis-restores-vision-in-blind-mice-and-detects-near-infrared-light-in-macaques.aspx>  
<https://www.asrs.org/patients/retinal-diseases/8/retinitis-pigmentosa-and-retinal-prosthesis>

<https://www.kaggle.com/code/gabrielloye/gru-vs-lstm-prediction>

| **Priority** | **Direction** | **Expected Deliverables** | **Unique Values Delivered** |
| --- | --- | --- | --- |
| 1 | Risk table by occupation subgroup | Stratified cognitive risk metrics and comparative risk tables | Actionable occupational risk stratification framework |
| 2 | Risk projection over 1, 5, 10 years | Time-based cognitive decline forecasts by subgroup | Personalized prognosis and healthcare planning tool |
| 3 | Work hours threshold identification | Thresholds for protective vs. adverse effects of work hours | Novel dose-response insights guiding occupational policies |

| **Step** | **Description** |
| --- | --- |
| Cluster occupation texts | Using KMeans, Agglomerative, or other clustering |
| Extract top terms per cluster | Use TF-IDF means or frequent words |
| Manually assign labels | Based on domain knowledge and top terms |
| Optionally refine with topic modeling | Use LDA/NMF to assist label generation |
| Map cluster IDs to labels | Create a dictionary mapping cluster number → label |