# Analyzing a Longitudinal Study on Cognitive Decline Using Mixed-Effects Models

Jeremy Faria

Wilfrid Laurier University

December 4th 2025

# Cognitive Decline

In Canada less than 1% of citizens ages 65-69 have dementia. This increases to 25% for ages 85 and older.

There are modifiable and non-modifiable factors that lead to cognitive decline. This research is aimed to:

- Encourage people to change their lifestyle
- Get tested sooner if they feel at risk

# Health and Retirement study (HRS)

- A multi-purpose longitudinal survey from 1992-2022 on United States citizens ages 50 and over

| Cohort Title | Birth Years | Year Added into Survey |
|---|---|---|
| HRS Cohort | 1931-1941 | 1992 |
| AHEAD Cohort | 1924 or earlier | 1993 |
| Child of Depression | 1924-1930 | 1998 |
| War Baby | 1942-1947 | 1998 |
| Early Baby Boomer | 1948-1953 | 2004 |
| Mid Baby Boomer | 1954-1959 | 2010 |
| Late Baby Boomer | 1960-1965 | 2016 |
| Early Generation X | 1966-1971 | 2022 |

Table: Cohorts of HRS Survey

# Survey Design

Step 1: Select primary sampling units (PSU)

↓

Step 2: Sample smaller areas in PSU's

↓

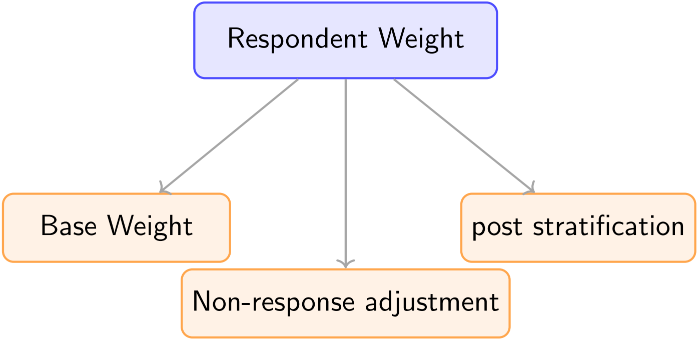
Step 3: A cluster of homes is selected

↓

Step 4: Age-eligible individuals are selected

# Survey Weights

Longitudinal analyzes use the latest available weight of the respondent.



# Cognition on the 27 point scale

The HRS used a clinician approved measurement on cognition using these tests:

- **10 points** for immediate noun recall
- **10 points** for delayed noun recall
- **5 points** for counting backwards from 100
- **2 points** for counting backwards from 20

Although not a formal diagnosis, clinicians believed the scores could be categorized into these groups:

- **12-27 points** - normal cognition
- **7-11 points** - cognitive impairment without dementia
- **0-6 points** - dementia

# Longitudinal Studies

- Able to track cognitive decline **overtime** through **repeated observations**
- Easier to find **cause and effect** relationships

# Covariance Structure

- $\vec{y} = [y_{11}, y_{1,2}, \ldots, y_{1,n}, y_{2,1}, \ldots, y_{m,n}]^T$ where $y_{ij}$ is observation j for subject i

$$var(\vec{y}) = \begin{bmatrix} V_1 & 0_{nxn} & 0_{nxn} & \cdots & 0_{nxn} \\ 0_{nxn} & V_2 & 0_{nxn} & \cdots & 0_{nxn} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{nxn} & 0_{nxn} & \cdots & \cdots & V_m \end{bmatrix}_{(m*n)x(m*n)}$$

$$\text{where } V_i = \begin{bmatrix} \sigma_{i1}^2 & \rho_{12}\sigma_{i1}\sigma_{i2} & \cdots & \rho_{1n}\sigma_{i1}\sigma_{in} \\ \rho_{12}\sigma_{i2}\sigma_{i1} & \sigma_{i2}^2 & \cdots & \rho_{2n}\sigma_{i2}\sigma_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\sigma_{in}\sigma_{i1} & \rho_{2n}\sigma_{in}\sigma_{i2} & \cdots & \sigma_{in}^2 \end{bmatrix}_{nxn}$$

- Violates the independence between observations assumption in linear regression

# Linear Mixed-Effects Model

For observation j of individual i the linear mixed-effects model can be written as:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \ldots + \beta_p x_{pij} + \gamma_{0i} + \gamma_{1i} z_{1ij} + \ldots + \gamma_{ki} z_{kij} + \epsilon_{ij}$$

- Fixed effects $\beta$ are population based inferences
- Random effects $\gamma_i$ are individual deviations from fixed effects for individual i

# Random Effects

$$\text{Random Effects}: \gamma_{0i} + \gamma_{1i} z_{1ij} + \ldots + \gamma_{ki} z_{kij}$$

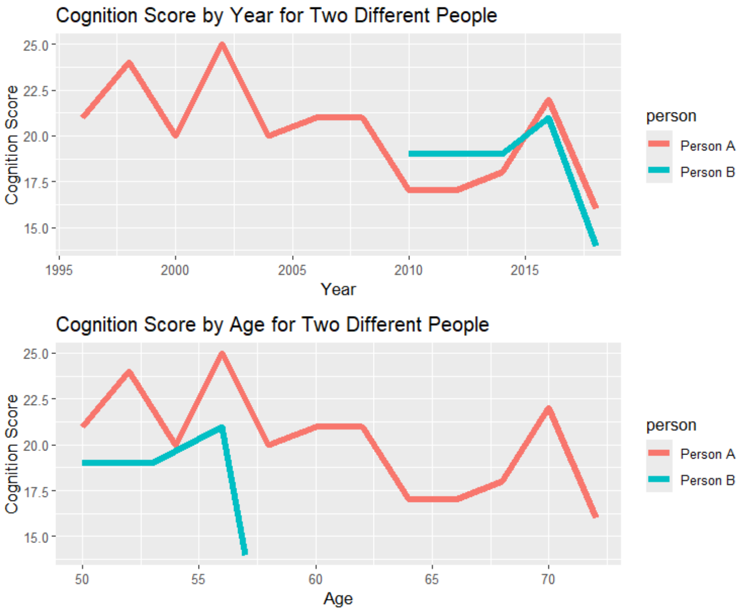
$$\gamma \sim MVN\left(\vec{0}_{kx1}, \begin{bmatrix} \sigma^2_{\gamma 0} & \sigma_{\gamma 0}\sigma_{\gamma 1} & \cdots & \cdots \\ \sigma_{\gamma 0}\sigma_{\gamma 1} & \sigma^2_{\gamma 1} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \sigma^2_{\gamma k} \end{bmatrix}\right)$$

- Random intercepts very common
- Random slope for time somewhat common
- Random slope for other covariates not as common

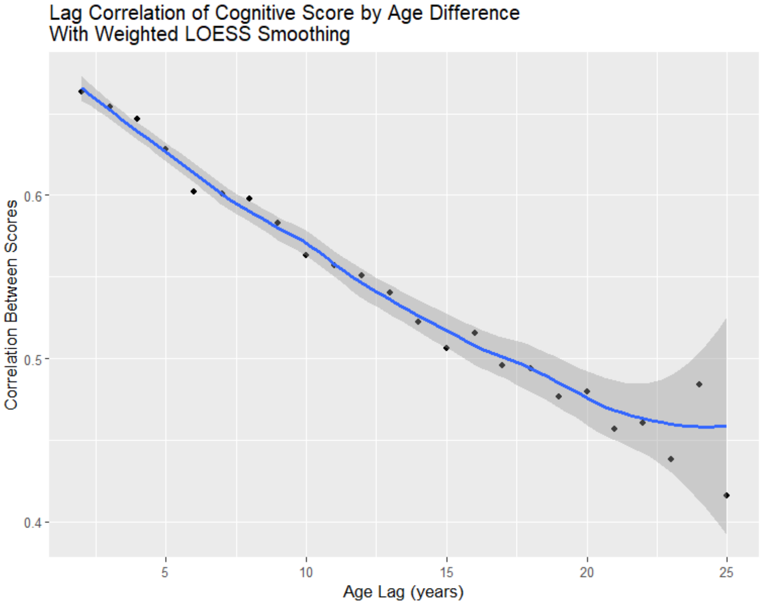
Random intercepts model correlation by:

$$Corr(y_{ij}, y_{ik}) = \frac{\sigma^2_{\gamma 0}}{\sigma^2_{\gamma 0} + \sigma^2}$$

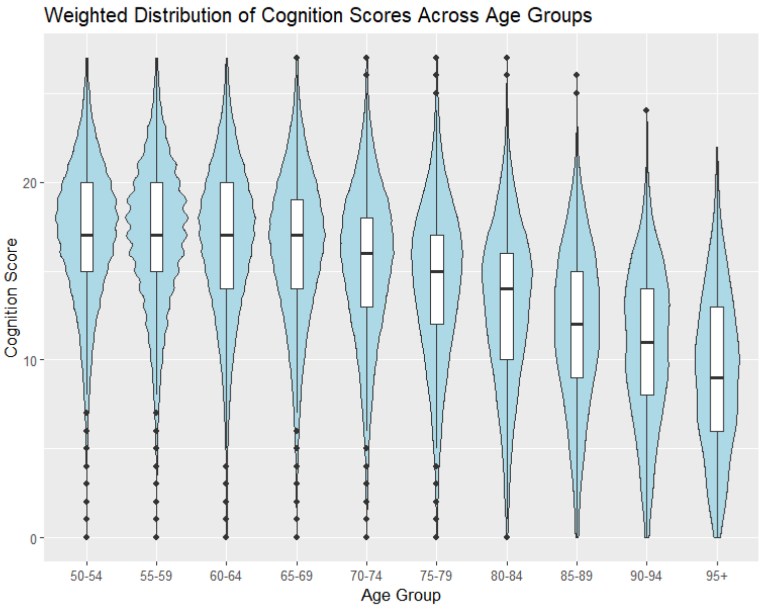
# Age as a Covariate



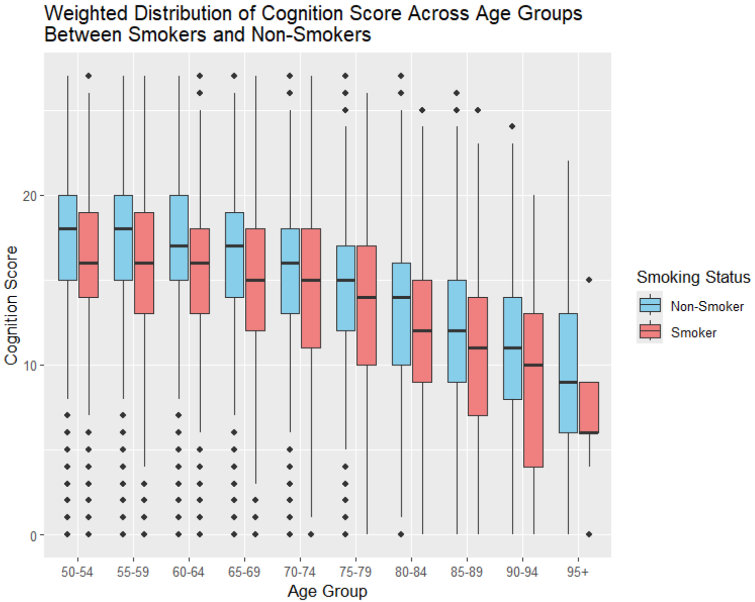
# Correlation Structure of HRS data



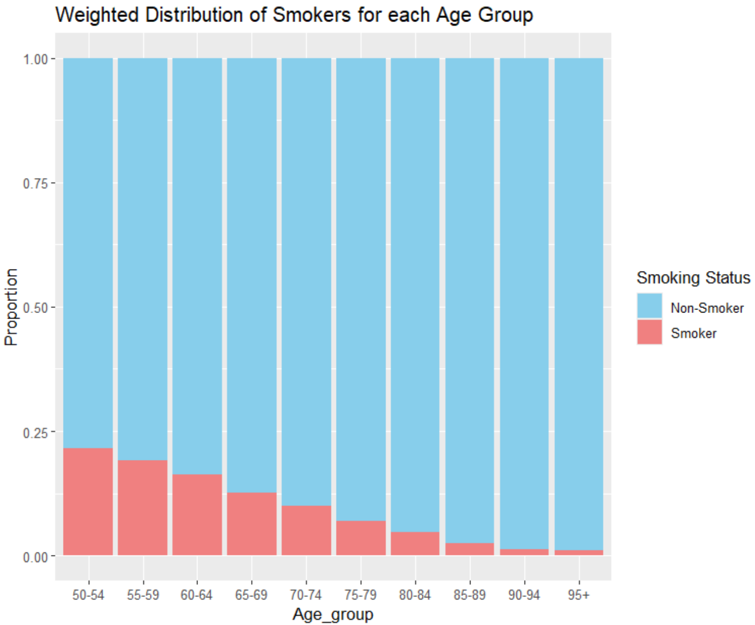
# Distribution of Cognition Scores in Age Groups



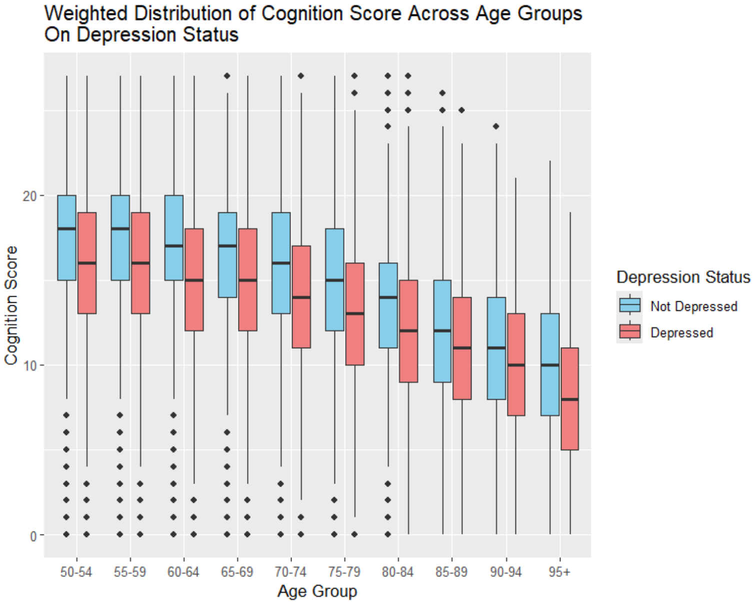
# Smoking as a Covariate



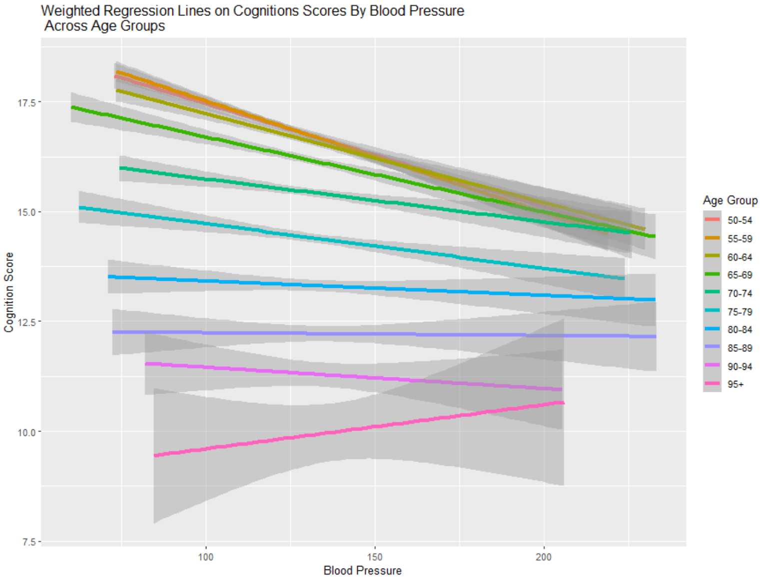
# Smoking as a Covariate



# Depression as a Covariate



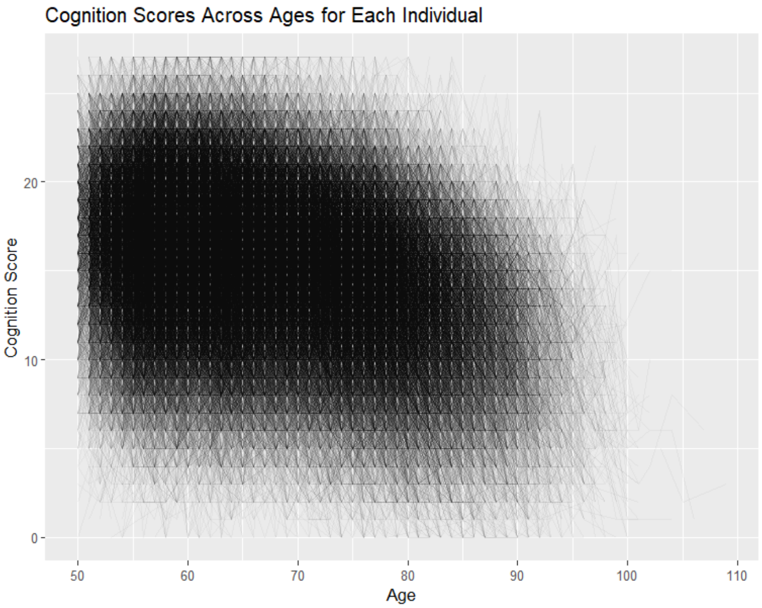
# Systolic Blood Pressure as a Covariate



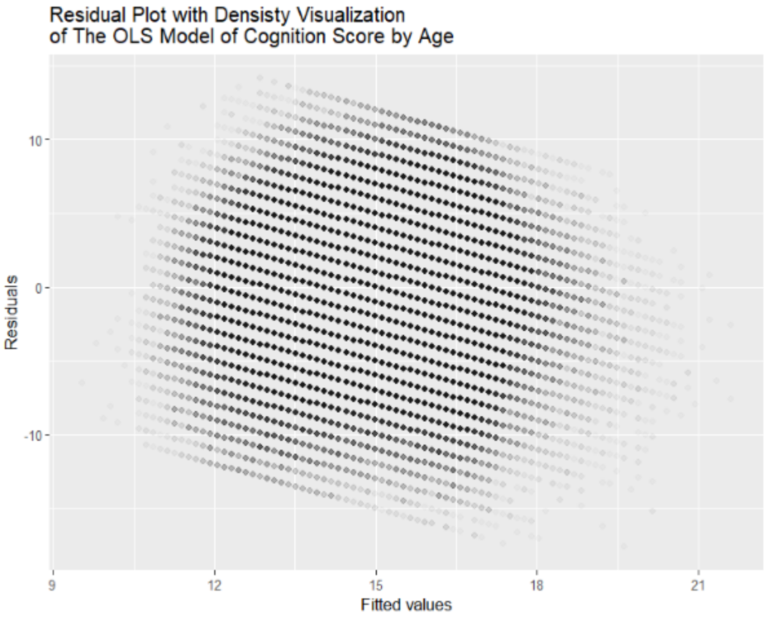
# Next Steps

- Set up models
- Estimation and hypothesis testing
- Model selection
- Diagnostics

# Appendix - Individual Cognition Scores Trajectories



# Appendix - Linear regression on Longitudinal Data



# Appendix - Restricted Maximum Likelihood Estimation

- Weighted Least Squares and Maximum Likelihood underestimate variance and covariance parameters
- REML provides unbiased estimates for variance and covariance

**MLE:**

$$\hat{\sigma}(V_0) = \frac{RSS(V_0)}{nm}$$

**REML:**

$$\hat{\sigma}(V_0) = \frac{RSS(V_0)}{nm - p}$$

# Appendix - Model Evaluation

- **Model comparisons:** AIC, BIC, Likelihood ratio test (random effects), marginal $R^2$ (fixed effects), conditional $R^2$

**Marginal:**

$$R_c^2 = \frac{Var(X\beta)}{Var(X\beta) + Var(Z\gamma) + Var(\epsilon)}$$

**Conditional:**

$$R_c^2 = \frac{Var(X\beta) + Var(Z\gamma)}{Var(X\beta) + Var(Z\gamma) + Var(\epsilon)}$$

- **Visually:** residual plots, violin plots of residuals of different times, QQ plots for residuals and random effects