**Workshop-Introduction to R**

**Basics in R**

Show how to load libraries, set up working directories, and import and export data in R.

**Data processing in R**

**Data:** paquid\_cog.csv and paquid\_cov.csv

* **Generate and label variables**

1. Generate a variable “fu”, which means follow-up time and equals to age-age\_init.
2. Generate a variable “dem\_young”, which means age of dementia onset (variable “agedem”) ≤70 years old (use the the if/else statement).
3. Rename variable “CEP” as “education” and change the variable class to factor. Label the variable values as 0=“Below primary school”, 1=“Primary school and above”.

* **Merge and reshape data sets**

1. Merge datasets “paquid\_cog” and “paquid\_cov” to a data frame named “paquid”.
2. Reshape the “paquid” data to wide format.

* **Row-wise calculation**

1. Generate a variable named “MMSE\_M”, which is the number of missing values across variables “MEM\_1”, “MEM\_2”, …, “MEM\_9” per individual. Label the variable as “the number of missing values in MMSE”.
2. View variables that contain “MMSE”.
3. Generate variables “MEM\_1”, “MEM\_2”, …, “MEM\_9”. which equals the mean of “BVRT” and “IST” at each time point.
4. View variables that contain “MEM”, “BVRT”, or “IST”.

* **Summarizing data**

1. Summarize variable “age\_init” (mean, sd, quantiles, etc), summarize “age\_init” by variable “male”.
2. Summarize variable “MMSE\_1” (mean, sd, quantiles, etc), summarize “age\_init” by variable “male”. Note how R deals with missing values.
3. Tabulate variable “male”, tabulate variable “male” and “education”, add row-wise and column-wise proportions.
4. Draw a histogram and a density plot of “MMSE\_1”.

* **Run simple models and check model output**

1. Run a linear regression, with “MMSE\_1” as dependent variable and “age\_init” and “male” as the independent variables, assuming “MMSE\_1” has a normal distribution. Check model output.
2. Run a logistic regression, with “dem\_young” as dependent variable and “male” as the independent variables.