At terminal …

~ Section 1~

Find individuals with a record of dementia, the diagnosis date, and individual's age.

Check that 41270 file contains all patients:

```
wc -l f.41270.tab 502505
```

Create a file containing just diagnosis codes of AD and check the number of patients:

```
grep -E 'F000|F001|F002|F009|G300|G301|G308|G309' f.41270.tab >
ad_lines_41270.tab
wc -l ad_lines_41270.tab 953
```

Create a file containing just diagnosis codes of VD and check the number of patients:

```
grep -E 'F010|F011|F012|F013|F018|F019|I673' f.41270.tab > vd_lines_41270.tab
wc -l vd_lines_41270.tab 507
```

Create a file containing just diagnosis codes of FTD and check the number of patients:

```
grep -E 'F020|G310' f.41270.tab > ftd_lines_41270.tab
wc -l ftd_lines_41270.tab 99
```

Create a file containing just diagnosis codes of OD and check the number of patients:

```
grep -E 'A810|F021|F022|F023|F024|F028|F051|F106|G311|G318' f.41270.tab >
od_lines_41270.tab
wc -l od_lines_41270.tab 555
```

Concatenate all files previous files creating a file containing ALLD and check the number of patients:

```
cat ad_lines_41270.tab vd_lines_41270.tab ftd_lines_41270.tab od_lines_41270.tab
> alld_lines_41270.tab
wc -l alld_lines_41270.tab 2114
```

Erase repeted lines, hence repeted ALLD patients:

```
sort alld_lines_41270.tab | uniq -u > alld_codes.tab
wc -l alld_codes.tab 1691
```

Create a file containing all eids codes of ALLD and check the number

```
awk '{print $2}' alld_codes.tab > alld_eids.tab
wc -l alld_eids.tab 1691
```

Check that 41280 file contains all patients:

```
wc -l f.41280.tab 502495
```

Create a file containing ALLD dates of diagnosis and check the number of codes:

```
grep -Ff alld_eids.tab f.41280.tab > alld_dates.tab
wc -l alld_dates.tab 1691
```

Create a file containing ALLD year of birth and check the number of codes:

```
grep -Ff eids_ALLD.txt 34.tab > alld_34.tab
wc -l alld_34.tab 1524
```

Create a file containing ALLD year of birth and check the number of codes: (2014 images)

```
grep -Ff eids_ALLD_2014.txt 34.tab > alld_34_2014.tab
```

```
wc -l alld_34_2014.tab 1135
```

~ Section 2 ~

Find individuals with a record of dementia, the diagnosis date, and individual's age.

Create files containing ALLD sex, age and assessment centre for analysis:

```
grep -Ff eids_ALLD.txt 31.tab > alld_31.tab

grep -Ff eids_ALLD.txt 34.tab > alld_34.tab

grep -Ff eids_ALLD.txt 54.tab > alld_54.tab
```

Take the desired ends from the full matrix (x_nan.csv) containing all individuals data in the whole cohort.

```
grep -Ff eids_ALLD.txt x_nan.csv > x_nan_for_dq.csv
wc -l x_nan_for_dq.csv 1524

grep -Fw -f eids_ALLD.txt x_nan.csv > x_nan_alld.csv
wc -l x_nan_alld.csv 1524

grep -Ff eids_ALLD.txt f.34.tab > f.alld_34.tab
wc -l f.alld_34.tab 1671

grep -Ff eids_ALLD.txt f.54.tab > f.alld_54.tab
wc -l f.alld_54.tab 1525
```

~ Section 3~

Healthy individuals selection

Create a new file with just the eids of healthy individuals:

```
awk '{print $1, $2}' f.41270.tab > reduced_41270.tab
wc -l reduced_41270.tab 502506

grep -E 'NA|f.eid' reduced_41270.tab > healthy.tab
wc -l healthy.tab 92189

awk '{print $1}' healthy.tab > healthy_eid.tab
wc -l healthy_eid 92189
```

Create a new file with all healthy individuals features:

```
grep -Fw -f healthy_eid.tab x_nan.csv > x_nan_healthy.csv
wc -l x_nan_healthy.csv 0
```

~ Section 4~

Create a new file with just the eids of healthy individuals:

```
grep -Ff healthy_eid.tab f.34.tab > f.healthy_34.tab
wc -l f.alld_34.tab 1671

grep -Ff healthy_eid.tab f.54.tab > f.healthy_54.tab
wc -l f.alld_54.tab 1525
```