# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

**Alterations of brain age in patients with schizophrenia**

ENIGMA-Schizophrenia Working Group

Version 0.4 – July 11, 2022

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

You will need four data files (described under Step 1) and the following R function files to run the analysis:

* get.means.R
* prepare.files.R
* model.fits.brainAge.R
* model.fits.brainAge.dx.R
* convert.PANSS.SANS.SAPS.R
* winsorize.R

You need to keep these files in your working directory. Please make sure that you have already done QA on the cortical and subcortical measures according to the previous ENIGMA protocols. If you need help with this, please contact us.

# Step 1: Prepare files

Please, make sure to have exactly the same header labels as shown in the example screenshots so that the commands used in this protocol need not to be changed!

Please see the header.xlsx file for a complete list of headers for all files.

All csv-files should be comma (“,” - without the quotes) delimited!

CorticalMeasuresENIGMA\_ThickAvg.csv, which contains your **cortical thickness** imaging phenotypes (after quality control) for all controls and schizophrenia patients.



CorticalMeasuresENIGMA\_SurfAvg.csv, which contains your **surface area** imaging phenotypes (after quality control) for all controls and schizophrenia patients.



SubcorticalMeasuresENIGMA\_VolAvg.csv, which contains your **subcortical volume** imaging phenotypes (after quality control) for all controls and schizophrenia patients.



Make sure that missing values and measures that were excluded from the analysis during QC are coded as “NA” without the quotes.

Covariates.csv, which contains the full set of variables we will be controlling for in each of the models. Please, make sure to have exactly the same header labels as shown in the example screenshot below (also see header.xlsx for a complete list). You will need to make this file. Missing values in the covariate columns are coded as “NA” without the quotes. The IDs in the SubjID column should match the format of the IDs in the SubjID column in the CorticalMeasuresENIGMA\_>brain.measure<.csv files.

In Excel, or your favorite text editor, create a series of columns with the following headers:

**Mandatory columns are:**

|  |  |
| --- | --- |
| **Header** | **Explanation** |
| SubjID | Subject IDs (same as in CorticalMeasuresENIGMA\_ThickAvg.csv) |
| Dx | Controls are coded as 0 and Patients are coded as 1 |
| Sex | Men are coded as 0 and Women are coded as 1 |
| Age | Age in years |

**Optional columns (for additional models and descriptives):**

If any of these are not available, include the header and add “NA” as data points. For the patient-related variables, please add “NA” for healthy control participants – do not code as “0” (=zero).

* **The Site column is** only required if your sample requires correction for data collection at multiple sites. If you have no site variables to include, do not include a Site column. NOTICE: we introduced a change, compared to previous ENIGMA-SCZ scripts. Please only include ONE Site column, using numbers to indicate, which participant came from which site. We will turn this into dummy variables in the script itself.

|  |  |
| --- | --- |
| Site | Only needed, if your data was measured at different locations / different scanners. Use only ONE site column, coding for n number of sites (see example covariate file below) |

* Measures of positive and / or negative symptoms (if you have participated in the negative/positive symptoms projects, you can use the measures from that project). If you don’t have a measure, include the column header and add “NA” (without quotation marks) as values for all individuals.

|  |  |
| --- | --- |
| SANS\_TOTAL | sum(of SANS items 1-7, 9-12, 14-16, 18-21, and 23-24) |
| SANS\_GLOBAL | sum(of SANS items 8, 13, 17, 22, and 25, which include affective flattening, alogia, avolition, anhedonia, and attention global rating scores, respectively) |
| PANSS\_NEG | negative PANSS score (sum of PANSS items 8-14) |

|  |  |
| --- | --- |
| SAPS\_TOTAL | sum(of SAPS items 1-6, 8-19, 21-24, and 26-33) |
| SAPS\_GLOBAL | sum(of SAPS items 7, 20, 25, and 34, which include hallucinations, delusions, bizarre behavior, and thought disorder global rating scores, respectively) |
| PANSS\_POS | Positive PANSS score (sum of PANSS items 1-7) |
| PANSS\_TOTAL | PANSS Total score or the sum of all thirty PANNS items |

* Additional measures:

|  |  |
| --- | --- |
| LENGTH\_OF\_ILLNESS | Length of illness in years |
| AO | Age of onset in years |
| IQ | Premorbid IQ |
| HAND | Handedness, coded as: 0 right; 1 left; 2 ambidexter |
| ETHNICITY | Ethnicity |
| PARENT\_SES | Parental socio-economic status |
| SUBTYPE | Current psychiatric disorder diagnosis;  For example, coded as:   |  |  | | --- | --- | | 295.1 | Disorganized type | | 295.2 | Catatonic type | | 295.3 | Paranoid type | | 295.4 | Schizophreniform disorder | | 295.6 | residual type | | 295.7 | Schizoaffectyive disorder | | 295.9 | Undifferentiated type | |
| MEDICATION | Medication, coded as: 1=unmedicated, 2=atypical, 3=typical and 4=both (and “NA” for healthy controls) |
| CPZ | Chlorpromazine equivalents |

Your covariate file should look similar to this (only comma-separated):



# Step 2: Obtain measures of brain age

* **Run the analysis in R:**

These scripts run best under R version 3.5. You can try and run them with an older version, but if you encounter any errors, you might want to update your R version and rerun.

The scripts require the following libraries:

"ppcor","lsmeans" ,"multcomp","data.table","plyr", “dplyr”, "ModelMetrics","caret","gridExtra","Hmisc","pastecs","psych","ggplot2"

You might want to install these beforehand, using:

install.packages("package\_name")

The script “1\_*protocol\_brainAge\_ENIGMA\_SZ\_20200306.R*” assumes that you are calling R from the same location as your four data files and all your R function scripts. If you are on Mac/Linux you can just cd to the directory on the command-line and then call R there. On a PC with Windows you can open R and use the setwd() command and give it the path to the folder that contains your files. For example, setwd("C:\\User\\Enigma\\") and then run the script by typing:

source("1\_*protocol\_brainAge\_ENIGMA\_SZ\_20200306.R*")

This script will pause at some point and produce the following output files:

* females\_raw.csv
* males\_raw.csv
* 1\_BrainAge\_SZ.log

You will need to upload the first two files (one at a time) to the PHOTON platform:

• <https://www.photon-ai.com/enigma_brainage>

These files only contain your imaging data, without the subject IDs. If you’re unable to upload

these files, please contact us for a local python-based solution.

Important acknowledgments:

If you use PHOTON and our ENIGMA FreeSurfer Brain Age Model for any of your own work, please make sure that you quote the following reference in any publications:

Han, L.K.M., Dinga, R., Hahn, T. *et al.* Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol Psychiatry* (2020). https://doi.org/10.1038/s41380-020-0754-0

Once you uploaded these files, you will receive (very quickly) an output file, which you need to

save and name as follows:

* females\_raw\_out.csv
* males\_raw\_out.csv

This output files looks just like your input file with an additional column in the end called “age\_prediction”.

# Step 3: QC and basic stats for your brainAge measure

In R, type: source("2\_*protocol\_brainAge\_ENIGMA\_SZ\_10082022.R*")

This will create four pdf files with figures, called:

* females\_BA\_descriptives.pdf
* males\_BA\_descriptives.pdf
* females\_Dx\_BA\_descriptives.pdf
* males\_Dx\_BA\_descriptives.pdf

# Step 4: Run brainAge SCZ analysis

In this step, we run the brainAge regression models.

This should output the following files:

* sumstats.bf.wins.Rdata (basic summary stats for all variables before winsorizing)
* sumstats.after.wins.Rdata (basic summary stats for all variables after winsorizing)
* several RData files started with “devAge\_”
* 2\_BrainAGE\_SZ.log (a log file; any errors will be written to this file)

Step 5: Run brainAge-ROI correlation analyses

In R, type: source("3\_*protocol\_brainAge\_ENIGMA\_SZ\_20210506.R*")

In this step, we run correlation analyses for brainAge and each FreeSurfer ROI.

This should produce the following files:

* ROI\_BA\_correlations.Rdata
* 3\_BrainAGE\_SZ.log (a log file; any errors will be written to this file)

----------