

Original version developed for ENIGMA-MDD
Hippocampal Subfield project (FS6.0dev). For
current studies based on full FS6.0 or FS7.1 please
refer to the alternative manual and script package.

Instructions part 1 (revised) and 2

ENIGMA-MDD hippocampal subfields project

12th of December 2016

***Designed for the Subfield module of the
developmental version of FS6.0***

Philipp Sämann, Neda Janashad,
Laura van Velzen, Derrek Hibar, Theo van Erp,
Christopher Whelan, Lianne Schmaal

Please use these instructions for the analysis of hippocampal subfield volumes in the MDD working group

If you have any questions or run into problems, please feel free to contact saemann@psych.mpg.de
or l.schmaal@ggzingeest.nl

These protocols are offered with an unlimited license and without warranty. However, if you find
these protocols useful in your research, please provide a link to the ENIGMA website in your work:
www.enigma.ini.usc.edu

Blueish parts in Courier Font can be directly used by copy & paste, with reddish
parts such as `<your_individual_path>` requiring editing according to your local directory
and file name structure. Nine files come along with this manual:

Name of the file	File description	Function
loop_all.sh	Shell script	start multiple segmentations
extract_subfields.sh	Shell script	extract subfield volumes
HippoColors_v6.0.txt	Plan text file	Color code table for subfields QC
outliers_hippo_and_QC_support.R ¹	R script	Histograms and QC support
QC_subfields_step_1_prepare_v2.sh	Shell script	Collect files needed for QC gallery
QC_subfields_step_1_prepare_old_v2.sh ²	Shell script	Collect files needed for QC gallery
QC_subfields_step_2_make_pngs_v2.m	Matlab script	Produce images for QC gallery
QC_subfields_step_3_webpage_v2.sh	Shell script	Arrange images in an *.html file
SUBFIELD_ANALYSIS_v1.6.R	R script	Perform actual statistical analysis ³

¹ This script is new – the functionality of the former script ENIGMA_HippoVolumes_plots.R (mainly generation of
histograms) is now covered by this script.

² Please use QC_subfields_step_1_prepare_old_v2.sh if your path to the result is [mri/]subfieldAnalysis_T1_v1.0/[...]

Step 1: Run hippocampal subfields segmentations using v6.0 in FreeSurfer

The project expects the subfield segmentation to be run by FreeSurfer v6.0. While FreeSurfer v6.0 is not released yet as a whole package (by 03/2016), the new hippocampal subfield segmentation algorithm v6.0 is already available within the developmental version of FreeSurfer. The download link can be found here:

<ftp://surfer.nmr.mgh.harvard.edu/pub/dist/freesurfer/dev>

General information and examples for subfield segmentations are given here:

<https://surfer.nmr.mgh.harvard.edu/fswiki/HippocampalSubfields>

An important application paper in this context that compared different FreeSurfer based subfield algorithms and explored heritability measures of hippocampal subfields can be found in:

<http://www.ncbi.nlm.nih.gov/pubmed/26747746>

NB: Note that subfields calculated in FS v5.3 cannot be included in this analysis due to changes in the boundary definitions and segmentation methodology. The prerequisite to run the FS v6.0 subfield segmentation is an accomplished standard segmentation in FS v5.3. Please contact us in the case you have not participated in that ENIGMA-MDD subcortical/cortical analyses and you wish support in running this general segmentation step, or if you have run the general segmentation by a version different from v5.3. The files that are needed in particular to continue with the subfield segmentation are: *aseg.mgz*, *nu.mgz* and *talairach.xfm*.

As soon as downloaded, the FreeSurfer environment needs to be defined (possibly, these basic settings are automatically called within the .cshrc file that is started when a new shell is opened, or any other automated startup script – in this case, skip these step):

```
csh
setenv FREESURFER_HOME <directory_containing_FS_installation>/freesurfer
source $FREESURFER_HOME/SetUpFreeSurfer.csh
```

In addition, the path to the subject folders needs to be defined:

```
setenv SUBJECTS_DIR <full_path_to_subject_folders>
```

The command `recon-all` should be automatically linked. Calling it without parameters will prompt an explanation of the function. Typing `which recon-all` should give you the location of the script in case you have several versions of FreeSurfer installed.

Calling the hippocampal subfield segmentation for a single subject:

³ Specific to the requirements of the ENIGMA-MDD hippocampal subfield project.

```
recon-all -s <subject_ID> -hippocampal-subfields-T1
```

Note 1: Segmentation from only T2-images and multimodal segmentation (more than 2 input channels) are generally supported by FreeSurfer. In this project we recommend to use only T1 images. Contact us in the case of questions hereon.

Note 2: The subfield tool requires the Matlab R2012 runtime software to be installed in the freesurfer folder. Please see footnote⁴ for details and install and unpack as required.

This starts the subfield segmentation script that usually takes about 30 min per case. The results are written into the already existing subfolder *mri*. There are 3 files produced per hemisphere, for the left hemisphere these file are:

- (a) The volumetric results listed in a file *lh.hippoSfVolumes-T1.v10.txt*, and
- (b) two images in mgz-format ref. as *lh.hippoSfLabels-T1.v10.mgz* and *lh.hippoSfLabels-T1.v10.1mm.mgz*.

Calling the hippocampal subfield segmentation for your complete sample requires a loop as follows:

1. Go to the FreeSurfer subject directory:

```
bash
cd <full_path_to_subject_folders>
```

2. Produce a list that contains the subjects for which you want to run the subfield segmentation tool, e. g. by:

```
ls -d1 <your_subject_folder_root>* > subjects.txt
```

Example: Your folders are labelled subject_001, subject_002 etc., then the command would be:

```
ls -d1 subject_* > subjects.txt
```

Double-check the list (no spaces, no other strings than subject folders etc.) by a standard text editor such as *jmacs* or *emacs* or *vi*.

3. Once checked, save the following lines as a script named *loop_all.sh* or use the respective script sent by mail:

```
#!/bin/bash
exec <subjects.txt
while read x; do
recon-all -s $x -hippocampal-subfields-T1
done
```

To make this script executable, type: `chmod a+x loop_all.sh`

Eventually, start the script by typing: `./loop_all.sh`

⁴ Download the runtime2012b.tar.gz package from the <https://surfer.nmr.mgh.harvard.edu/fswiki/MatlabRuntime> to your Freesurfer folder ([.....]/freesurfer) and unpack it by typing `tar xvf runtime2012b.tar.gz`

Step 2: Collect the subfield volumes into a file *hippo_subfields.csv*

For this, a script *extract_subfields.sh* has been distributed that collects all subregional volumes and the ICV into a file called *hippo_subfields.csv*. Make sure the script is executable by typing

```
chmod u+x extract_subfields.sh
```

and copy the script to the your FreeSurfer subject directory. To make it loop through the right subject folders, there are two ways to customize the script:

Way A: Change the line

```
for subj_id in $(ls -d WG_*); do
```

to a name fitting your subject folders, e. g. subject_001, subject_002, would result in:

```
for subj_id in $(ls -d subject_*); do
```

Way B: Use a subject list from a file, e. g. */home/user/somefolder/subjects.txt*

Then, in the script the line

```
for subj_id in $(ls -d WG_*); do
```

would then need to be changed to:

```
for subj_id in $(cat /home/user/somefolder/subjects.txt); do
```

(Ideally you use the same *subjects.txt* file already defined during step 1.)

Eventually, run the script:

```
./extract_subfields.sh
```

In addition to the left and right hippocampal subfield volumes, the left and right hippocampal volume (as used for ENIGMa-MDD) and 3 global values (eTIV, brain volume, total GM) are extracted.

The result is a comma-separated (CSV) file that can be opened in your favorite spreadsheet application (e. g. Excel) or a text editor (e. g. Emacs). The first row is a header describing the extracted regions and names for each column. Each row after the first gives the volume measures for each subject found in your FreeSurfer directory. In the next step, you will do a QC of the segmentation quality.

Note 1: if you get error messages such as:

```
extract_subfields.sh: line 13: printf: 491.769324: invalid number
```

run the following command before running the `sh extract_subfields.sh` command:

```
export LC_NUMERIC="en_US.UTF-8"
```

Note 2: When you edit the files in Excel, be sure to keep them in CSV format when you save and check whether the columns are still separated by commas and text strings are *not* enclosed by inverted commas.

Step 3: Outlier detection, histogram generation and QC support

This step requires that you have R installed and that you have saved the script *outliers_hippo_and_QC_support.R* in the same directory as your *hippo_subfields.csv* file.

1. Change directories to the location of your *hippo_subfields.csv* file generated in step 2.
`cd <full_path_to_hippo_subfields.csv_file>`
2. Make sure the *outliers_hippo_and_QC_support.R* script is also in that directory and run:
`R --no-save --slave < outliers_hippo_and_QC_support.R > QCsupport.log`

This will generate a file *QCsupport.log* that contains the result of two type of 'abnormality' checks:

- Outliers across the group: both the subject and the affected subfield or global variable are listed.
- 'Rank violations': Exploration of large datasets demonstrated that subfield volumes show a certain ranking within a subjects. Here, both too low hippocampal tail volume and subiculum not being ranked #3 are considered as suspicious for segmentation failures.

As both tests may lead to the same cases, a list of all subjects that need specific QC is printed at the end of the log-file. *Example:*

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Resulting minimum list of cases fom steps 1 & 2 for visual QC %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Perform QC on WG_050
Perform QC on WG_056
Perform QC on WG_075
Perform QC on WG_078
[...]
```

See step 6 how to perform the visual QC on these cases.

Pack together the resulting **histograms** and **summary statistics**

```
tar -cf <YourSampleName>_SF_QC.tar *_hist_volumes.png SummaryStats_hippo_subfields.txt
```

Mail the resulting **tar* file to: saemann@psych.mpg.de

NB: Just because a subject is an outlier or because it deviates from the average rank order does not necessarily mean they should be excluded from the analysis. The outlier detection should merely turn your attention to such cases, as extreme values may be a hint towards technical failure of the segmentation.

Step 4: Preparing the visual QC

Visual Q/C means that images with an overlay of the hippocampal subfields on the T1 are viewed and assessed. As the subfield segmentation is based on the general delineation of the hippocampus during the general subcortical segmentation as part of the volume-based processing stream called by *recon-all*, the subfields usually (>98%) lie correctly in the hippocampal area.

Three scripts (steps 4a to 5c) have to be run to prepare the actual visual Q/C.

Note: In case you have outliers in the distribution of one of the 3 global variables (eTIV, Brain, TotalGM), this requires going back to the Q/C of the segmentation of the cortical surfaces and cortical ribbons using the scripts used for the ENIGMA-MDD cortical project. If you have not participated in the ENIGMA-MDD cortical project and need support in this step, please contact me (E-Mail: saemann@psych.mpg.de).

Step 4a: Preparation step: Shell script *QC_subfields_step_1_prepare_v2.sh*

This step creates a folder *assembly* within the already existing FreeSurfer folder *mri*. Copies of a background image (nu.mgz), the general segmentation result file (aparc+aseg.mgz) and the hippocampal subfields ([r/l]h.hippoSfLabels-T1.v10.mgz)⁵ are put into this folder, images are transformed to NIFTI format and sampled to a resolution suitable for overlay. Before running the script, three adaptations within the *QC_subfields_step_1_prepare_v2.sh* need to be made:

- Adaptation 1: Define the variable ***subject_folder*** as needed.
- Adaptation 2: Define the location of FreeSurfer/bin (i. e. variable ***FS_bin***).
- Adaptation 3: Define the list of subjects, i. e. change the line
`for subject_id in $(cat /data2/remake_enigma-mdd/N550/done_subfield/list_folders.txt); do`
to match the location and name of your subject list (ideally, again *subjects.txt* as defined).

Then make sure the script is executable

```
chmod a+x QC_subfields_step_1_prepare_v2.sh
```

and run it (from any location) by calling:

```
./QC_subfields_step_1_prepare_v2.sh
```

⁵ Use the script *QC_subfields_step_1_prepare_old_v2.sh* if your path to the result is [mri/]subfieldAnalysis_T1_v1.0/[...]

Step 4b: Create overlay images: Matlab script *QC_subfields_step_2_make_pngs_v2.m*

For this step you need Matlab and the (free) set of help functions to handle NIFTI images:

<http://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image>

Make sure that the Matlab command *load_nii* is linked and can be called.

Then open the distributed script *QC_subfields_step_2_make_pngs_v2.m* in the Matlab editor and change the 6 variables in the area “ADAPT AS REQUIRED” as follows:

Variable	Comment
subject_folder	Full link to your FS subject folder
dirOUT_general	Link to an (empty) folder into which the Q/C images will be written out.
HIPPOmap	Link to the distributed color-code table <i>HippoColors_v6.0.txt</i>
fid	Change this line so it matches the link to your subject file.
columnwidth	This is for a case that your subject list contains folder names of different length (e. g. “subject_sample_AGHG_01” next “to subject_X1”). If unsure, make a guess of the longest string and add +10...
transparency	This defines, how transparent the subfields are when overlaid onto the T1-image [0=transparent, 1=opaque]. A value of 0.8 is a good starting point.
transparency_aseg	Transparency of overlay of recon-all based hippocampal mask; default is set to 0.15.

After these adaptations, run the Matlab script. It will take about 20-40 seconds per case, yet, note, that all cases are processed twice (first pass: classical output with colored subfields; second pass: T1WI plus overlay with aseg-result and hippocampal fissure, and T1-WI only). The result should be several *.png files in the respective subject subfolders.

Step 4c: Collect all *.png images into a browsable html-file

The step pulls together all *.png files into a html file that can then be loaded into any standard browser. Please adapt the 'WG_' in the respective line in the *QC_subfields_step_3_webpage_v2.sh* script: `subject_list='ls -d1 WG_*` to match your subject folders' names.

If you want to reduce the amount of cases merged in the html-file (as some browsers crash in the case of too many subjects), there are two simple solutions:

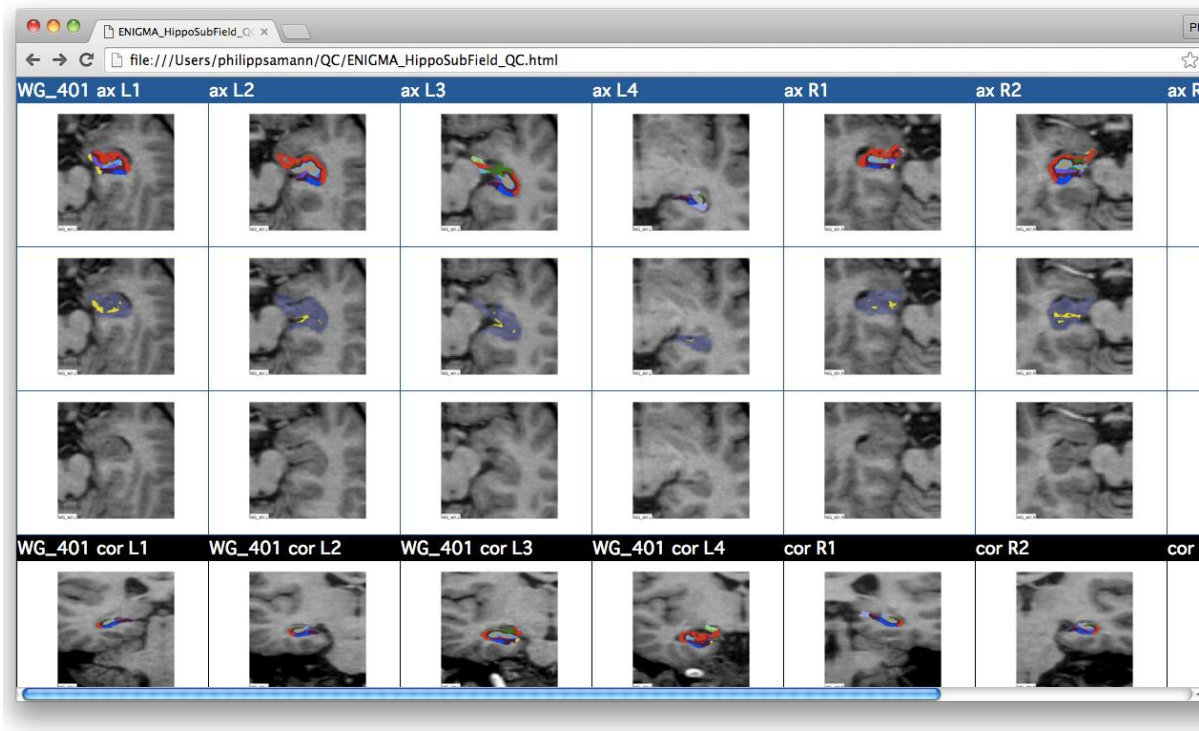
- (1) Move some cases temporarily away into a subfolder,
- (2) restrict the selection for the subject_list,

When you call the script, the Q/C folder must be given as argument:

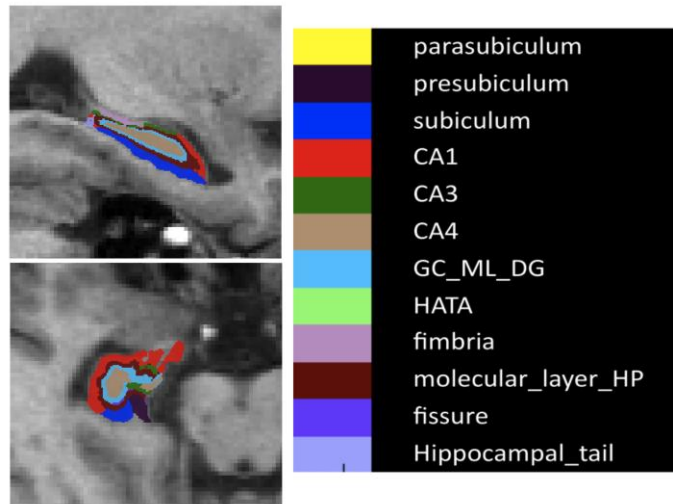
```
./QC_subfields_step_3_webpage_v2.sh <full_path_to_QC_folder>
```

Note: The folder likely corresponds to the variable `dirOUT_general` in the previous Matlab script.

The result is a browsable file `ENIGMA_HippoSubField_QC.html` that can be loaded into standard browser such as *Firefox* or *Explorer* or *Chrome* (recent versions recommended; in addition, working memory overload can be avoided by closing all other applications before). The scheme is as follows: 2 x 12 axial, 2 x 12 coronal and 2 x 12 sagittal images. The images can be enlarged by clicking on them.

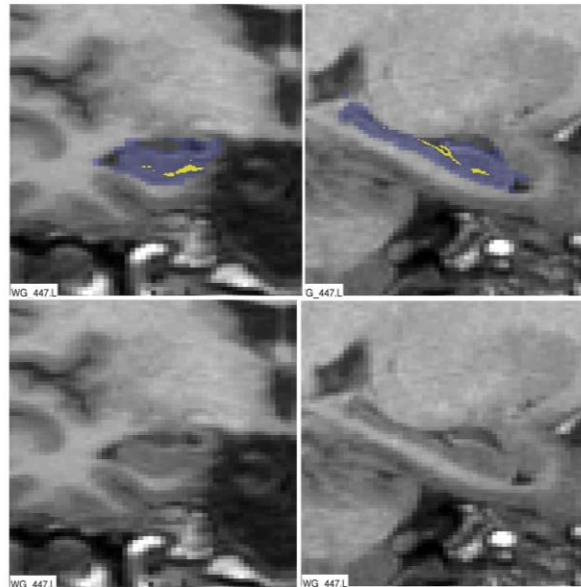


The **first** row shows the subfields with the color codes as listed on the right. Note that the hippocampal fissure is also included (purple).



The **second row** shows an overlay of the T1 with
 (a) the **whole hippocampus** as returned from the general subcortical segmentation (aparc+aseg.mgz file), depicted in blueish/transparent,
 (b) the **hippocampal fissure** depicted in yellow.

The **third row** shows the (bias-corrected) T1 without any overlay.

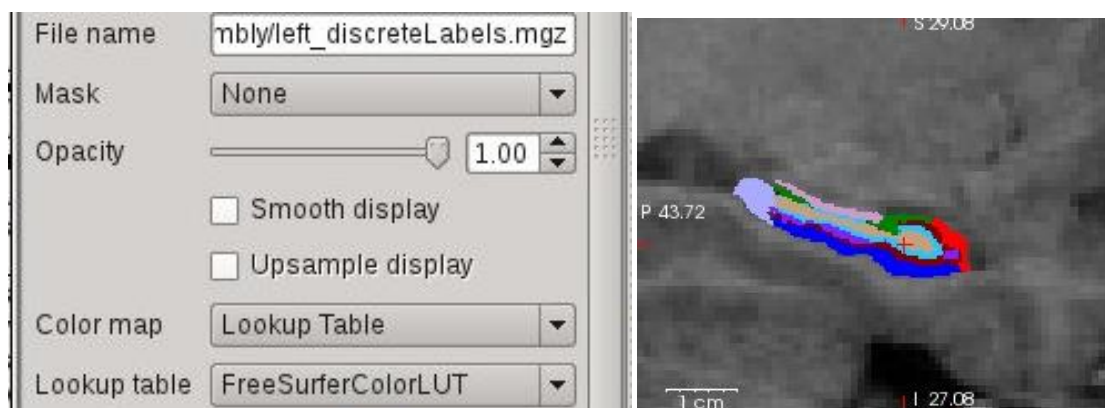


Note: In addition to browsing the (static) html-file, for a single case, one should also view at least about 10 cases dynamically by calling the FreeSurfer viewer within the subfolder *assembly*:

`freeview nu.gz left_discreteLabels.mgz` (for the left hemisphere)

`freeview nu.gz right_discreteLabels.mgz` (for the right hemisphere)

If you switch the Color map to *Lookup Table* and select *FreeSurferColorLUT*, the same colors as in the html-file should be displayed in FreeSurfer. It is recommended to scroll through the cases to get an impression of the variability of the subfields and their appearance at different slice positions:



Hints for the visual QC

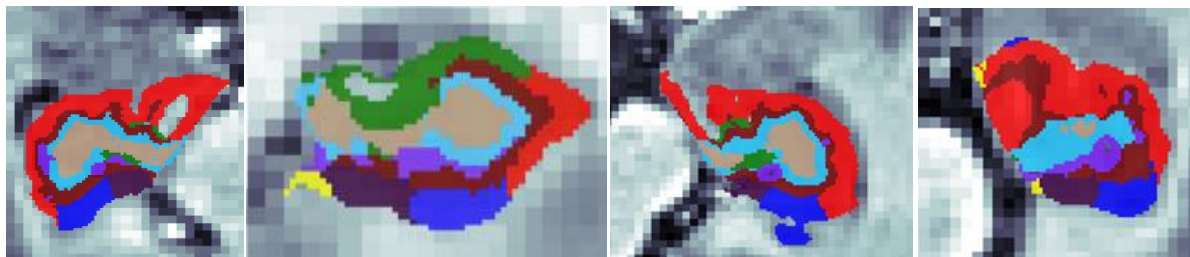
The goal of the visual QC is to (i) get an overview of the segmentation quality of the whole sample, and (ii) to identify cases that should be excluded as a whole. We currently do not recommend to exclude single subfields of one case.

1. The general segmentation of the subcortical structures (the FreeSurfer volume processing pipeline) has probably undergone QC during previous projects, based on the standard ENIGMA protocols⁶. If not, the hippocampal segmentation result is now again displayed in the respective **2nd row** in the browser. Here, no larger portions of the hippocampus should be cut off, and the mask should not be shifted (by several millimeters) as sometimes seen in completely failed cases.

2. **Which cases should undergo visual QC?** Ideally, all cases are briefly checked for good positioning of the hippocampal mask as a whole, and fissure within the hippocampus. As minimum, cases list after executing the R script *outliers_and_QCsupport.R* should be reviewed.

3. The **hippocampal fissure** is generated by the segmentation algorithm, yet does not represent a proper tissue subfield but is a thin CSF space within in the complex hippocampal convolute. On the sagittal slices and most coronal slices, the fissure is typically positioned *within* the hippocampus mask. By assessing the fissure location one avoids to circularly use the subfields that are later used as variables of interest in the analysis. The corresponding **T1WI slices** (without overlay) are displayed in the 3rd row. Finding the fissure within the hippocampus (ideally in areas with corresponding low T1 signal) with its typical longish configuration on sagittal slices is a good indicator that the probability subfields atlas was properly coregistered with the T1WI.

4. **Should single subfields be excluded?** No, rather the whole subject is excluded in the case of severe segmentation problems. Technically, it is sufficient to set *one* subfield in the *hippo_subfields.csv* table to NA; the scripts will interpret this as complete exclusion. Below are four examples from typically varying samples that need not necessarily be removed (from left to right): (1) "Hole" in CA1 (red), (2) "hole" in CA3 (green) (caused by partial volume effects in relation with the alveus, a white matter structure not written out); (3) extension of the subiculum (blue) with neglectable volume effect), and (4) bulky CA.



If you would like to discuss individual cases, the QC folder can be compressed and uploaded onto the MPIP ftp-server – access is provided after a short mail to saemann@psych.mpg.de

⁶ <http://enigma.ini.usc.edu/protocols/imaging-protocols/quality-checking-subcortical-structures>

Step 5: Exclude poor segmentations and pack files

Based on visual QC, you may have identified cases with poor segmentations or starting locations. You can try to re-run

```
recon-all -s <SubjID_to_be_rerun> -hippocampal-subfields-T1
```

for those cases and go through step 2–4 again (however, usually, there is high reproducibility in that the new result will be very similar). If the result remains suspicious, exclude these volumes from *hippo_subfields.csv* by marking **any subfield with NA**. Any values replaced by NA in any of the subfields (or the global volumes) will lead to exclusion of the whole case from histogram generation, the summary stats and the actual later main analysis. Therefore, it is important to re-run step 3 after marking cases with NA.

Step 6: Main analysis

*The following instructions provide you with a wide set of first level results useful for a subsequent meta-analysis. The prerequisite is a covariates file constructed exactly as specified in ***syntax.pdf***.*

The actual statistical analysis needs two input files:

1. *hippo_subfields.csv*

2. *Covariates.csv*

These input files need to be located in the folder from which you start R.

Then, call R and then check that the *ppcor* package is installed. Check by the command

`search()` – if the *ppcor* library is not listed, install it by:

```
install.packages("ppcor")
```

and activate the library by

```
library("ppcor")
```

Eventually, run the analysis script by:

```
source("SUBFIELD_ANALYSIS_v1.6.R")
```

Three variables can flexibly be altered by the user, yet, it should first be run with the default settings:

whole	defines the global correction variable; default is 'eTIV'; theoretical alternatives are 'Brain' or 'TotalGM'
write_logfile	default = 1: 0 = report messages on the fly; 1 = write logfile
AgeSq_and_interactions	default = 1, so age ² , gender-by-age and gender-by-age ² are added to (mostly) all models; 0 = the three terms are excluded from the models

The output will be many **Rdata* files and the **log*-file (if requested). Please pack these files together using

```
tar -cf *Rdata *log <YourSampleName>_SF_ANALYSIS.tar
```

```
gzip <YourSampleName>_SF_ANALYSIS.tar
```

These files were then collected from the different sites and used for the ENIGMA-MDD Hippocampal Subfields meta-analysis. In case you are interested in such analysis and you would like to use existing R scripts for this, please contact Dr. med. P. G. Sämann (saemann@psych.mpg.de) or Dr. L. Schmaal (lianne.schmaal@unimelb.ed.au) for details.