# ENIGMA AN DTI processing, TBSS and ROI analysis guidelines

[ENIGMA AN DTI processing, TBSS and ROI analysis guidelines 1](#_Toc135669414)

[Aims 2](#_Toc135669415)

[File Structure 2](#_Toc135669416)

[Preprocessing 3](#_Toc135669417)

[Getting the average B0 mask 4](#_Toc135669418)

[Calculating tensor metrics 4](#_Toc135669419)

[Image QC 5](#_Toc135669420)

[TBSS 5](#_Toc135669421)

[Python versions 6](#_Toc135669422)

[Installing TBSS 6](#_Toc135669423)

[Running the TBSS pipeline 6](#_Toc135669424)

[Making the image list and subject list files 7](#_Toc135669425)

[QC 7](#_Toc135669426)

[Bad registration (skip this section if registration is fine) 7](#_Toc135669427)

[Analysis 8](#_Toc135669428)

[Making the covariates file 8](#_Toc135669429)

[Running the R script 10](#_Toc135669430)

[Local results 11](#_Toc135669431)

## Aims

The aim of the TBSS analysis is to describe white matter alterations that are present among individuals with AN relative to healthy control peers (HC). This is possible via analysis of diffusion imaging data, which is sensitive to differences in the direction of water diffusion throughout the brain. When diffusion is more directed within white-matter tracts, it is assumed these tracts have better structural integrity. Where diffusion is more isotropic within white-matter (i.e., undirected) this can suggest poorer integrity of the brain tissue, potentially due to axonal degeneration or reduced myelination/axon density. In this analysis we will estimate the direction of water flow in each voxel along a white matter skeleton. Average diffusion metrics (across voxels within 25 distinct regions of interest) will be calculated, and compared between individuals with AN and HC. The between-group analyses will generate a set of test statistics that will be combined across different sites in a meta-analysis, for large-scale comparison of white matter characteristics between individuals with AN and HC.

These instructions are intended to allow the different sites to complete the analyses and submit results for the meta-analysis, however **the processing and analyses can be completed centrally if data sharing agreements permit this**.

## File Structure

Before starting the analysis, make sure that you have downloaded the github repository. Navigate to <https://github.com/CaitlinLloyd/ENIGMA_AN_DTI>, and select the code button, and download zip.

Graphical user interface, application

Description automatically generated.

Save the folder to a location you have access to, and rename the Project\_name folder to the name of your project. **In this example, the project name is NYSPI.** Next, move your dwi data to the Data folder inside the ENIGMA\_AN\_DTI/project\_name folder. Please structure your MRI data according to the BIDS structure (see example below). This will ensure that all the scripts work as intended. As we proceed through the instructions, pay attention to the folder names, as the scripts are designed to work on these names – changing the folder names could lead to the code not working.

Background pattern

Description automatically generated

Throughout the document you may see the $ sign followed by a generic name. This is simply indicating a variable. For example, we may use $sub to refer to a subject id that changes – in the code you would replace $sub with the actual subject id (e.g., sub-101).

## Preprocessing

**You can ignore this step if preprocessing is already completed.**

DWI data require preprocessing prior to estimating the diffusion tensor and completing the TBSS analysis. The pipeline ideally includes the following steps:

* Denoising
* Susceptibility distortion correction
* Motion correction
* Eddy current correction

Notably, deringing (removal of Gibbs ringing artifact) is optional since there is little evidence this step is necessary.

Prior to preprocessing, check your data for artifacts. This resource might be helpful:

https://sites.google.com/a/labsolver.org/dsi-studio/course/diffusion-mri-signals

In the event that artifacts are observed, you can remove them using the remove\_badvols script (on the project github page), which will drop the volumes from both the dwi series and the bval and bvec files.

Preprocessing methods are not prescribed (you are free to use whichever programmes you would like providing they include the recommended steps). Please direct any queries to [Caitlin.lloyd@nyspi.columbia.edu](mailto:Caitlin.lloyd@nyspi.columbia.edu) if you would like to check any aspects of your preprocessing pipeline. We do recommend using QSIprep however, since you can complete all of the steps using a convenient wrapper. QSIprep installation information is available here: <https://qsiprep.readthedocs.io/en/latest/installation.html>

Once installed, the following command will ensure all corrections noted above are implemented:

qsiprep -i $inputdir -o $outputdir -d $workdir -l qsilog.log --output-resolution 1.00 --separate-all-dwis --participant\_label $subname --hmc\_model eddy --fs-license-file $FS\_license\_location

If there are GPUs available on your computing system, you can implement slice to volume correction using FSL eddy. In this case, the command looks a little different, and includes the eddy\_config flag:

qsiprep -i $inputdir -o $outputdir -d $workdir -l qsilog.log --output-resolution 1.00 --separate-all-dwis --participant\_label $subname --hmc\_model eddy --fs-license-file $FS\_license\_location eddy\_config eddy\_params.json

An example of the eddy\_params.json content is on the Github repository (within the preprocessing folder). You will need to provide your own slice order information (in a file called slspec.txt) and specify mporder after following the guidelines provided by FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/eddy/UsersGuide#A--mporder) to identify a reasonable parameter.

## Getting the average B0 mask

To complete FSL’s dtifit that allows for calculating FA, MD, RD, AD, we we will need a brain mask extracted from the average of the preprocessed B0 images. To do this we can use tools from [FSL](#FSL) or [mrtrix](#mrtrix) (<https://www.mrtrix.org/>).

***In the step below, either follow the FSL or the MRtrix option:***

**If you want to use FSL, you can use the following to extract and average the B0 images:**

B0\_extract $bval\_file $preprocessed\_dwi\_image $meanB0image

To produce BIDS compliant files, which will be useful in the next stage, the command would look something like this:

B0\_extract sub-0101\_ses-001\_dwi\_preprocessed.nii.gz sub-0101\_ses-001\_dwi\_preprocessed\_averageb0.nii.gz

**Using mrtrix, the commands to identify the b0 volumes and average across them are:**

mrconvert $preprocessed\_dwi\_image.nii.gz $preprocessed\_dwi\_image.mif -fslgrad $bvec\_file $bval\_file -datatype float32 -stride 0,0,0,1

dwiextract $preprocessed\_dwi\_image.mif --bzero -quiet -nthreads 4 | mrmath - mean -force $meanB0image -axis 3 -quiet -nthreads 4

To produce BIDS compliant files, which will be useful in the next stage, the commands would look something like this:

mrconvert sub-101\_ses-001\_dwi\_preprocessed.nii.gz sub-101\_ses-001\_dwi\_preprocessed.mif -fslgrad sub-101\_ses-001\_dwi.bvec sub-101\_ses-001\_dwi.bval -datatype float32 -stride 0,0,0,1

dwiextract sub-101\_ses-001\_dwi\_preprocessed.mif --bzero -quiet -nthreads 4 | mrmath - mean -force sub-101\_ses-001\_dwi\_preprocessed\_averageb0.nii.gz -axis 3 -quiet -nthreads 4

To extract the B0 brain mask (after obtaining the B0 images), use the following:

bet2 $meanB0image $meanB0image\_mask -m -f 0.2 -v

To produce BIDS compliant files, the commands would look something like this:

bet2 sub-001\_ses-001\_dwi\_preprocessed\_averageb0.nii.gz sub-001\_ses-001\_dwi\_preprocessed\_averageb0mask.nii.gz -m -f 0.2 -v

## Calculating tensor metrics

The diffusion tensor is a three-dimensional model of the direction of water flow, and is described using eigenvectors and eigenvalues that indicate the direction and magnitude of water diffusion along three axes, in each voxel. Using the diffusion tensor, several metrics describing the direction of water flow can be calculated, and used to make inferences about the integrity of white matter throughout the brain. These voxel-wise metrics are represented in 3D brain volume maps, and include the following:

Mean diffusivity (MD): Average amount of apparent diffusion along each of the three diffusion axes.

Axial diffusivity (AD): The amount of apparent diffusion along the principal diffusion axis.

Radial Diffusivity (RD): The average amount of apparent diffusion along the secondary and tertiary diffusion axes.

Fractional anisotropy (FA): Describes the extent of water diffusion along the principal direction relative to the orthogonal directions (a value of 0 indicates the absence of directed water movement; a value of 1 indicates water movement in a single direction only).

Running the command below will allow you to fit the diffusion tensor to each voxel, to calculate metrics describing the quality of the white matter tracts that will be compared between patient and control groups. This step will need to be completed for each participant, and could be run in a loop. Where and how?

dtifit –data=$preprocessed\_dwi\_image –mask=$averageb0mask –bvecs=$ dwi.bvec –bvals=$dwi.bval –out=$outputfile

RD is the first eigenvalue (\_L1 image), and AD can be calculated by averaging the second and third eigenvalue map files (\_L2 and \_L3 files). You will need to rename the L1 files for every subject, and average the L2 and L3 files manually:

mv $L1.rii.gz $RD.rii.gz

fslmaths $sub/${sub}\_dtifit\_L3.nii.gz -add $sub/${sub}\_dtifit\_L2.nii.gz -div 2 $sub/${sub}\_dtifit\_AD.nii.gz

The bash script on github (tensor\_metrics) will perform the tensor metric calculation in a loop. To use this script you will need to type the following into the command line:

./tensor\_metrics $indir

For this to work, the files should be named as follows:

Preprocessed file: ${sub}\_ses-001\_dwi\_preprocessed.nii.gz

Mask: ${sub}\_ses-001\_dwi\_averageb0mask.nii.gz

Where $sub is the subject ID (e.g., sub-101)

Indir is the ENIGMA\_AN\_DTI folder.

Image QC  
To make sure the tensor images are appropriate to use in the analysis, we check whether the principle vectors computed during DTIfit are aligned to the white matter tracts. You can do this in any viewing software (see <https://fsl.fmrib.ox.ac.uk/fslcourse/2019_Beijing/lectures/FDT/fdt1.html#dtifit>, DTI output section), however the ENIGMA DTI working group have written some helpful scripts for creating a QC webpage for your data. This does rely on matlab, and instructions are available here: <https://git.ini.usc.edu/ehaddad/03_enigma-dti-quality-control>

## TBSS

All of the steps required for the tbss analysis can be completed quickly and easily using the TBSS repository developed by Psychiatry Neuroimaging Lab @ BWH/HMS (<https://github.com/CaitlinLloyd/TBSS_ENIGMA_AN> ). We are using a slight adaption of the original repository to enable the acquisition of additional metrics. The commands outlined below work well on a linux/unix machine. If you are using a different operating system (e.g., Windows), and you are running into errors, please contact [caitlin.lloyd@nyspi.columbia.edu](mailto:caitlin.lloyd@nyspi.columbia.edu).

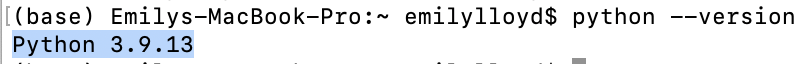
The pipeline produces a series of metrics that describe the properties of the major white matter tracts of the brain. The inputs are the diffusion tensor metric maps (i.e., FA, MD, AD, RD). The program erodes the FA maps slightly, and then performs a linear registration of the participant FA images to a FA template (here the ENIGMA reference image, which was created from aligning the FA maps of 100 healthy participants). This transform is also used to warp the other maps (i.e., MD, AD, RD) into ENIGMA space.

The ENIGMA skeleton is derived from the mean of the FA images from this representative sample, having identified the fibre bundles that are common to all participants. White matter skeleton images for each participant are created by filling the ENIGMA skeleton with FA values from the nearest relevant tract centre, for each individual. This is repeated for other modalities (i.e., MD, AD, RD). The skeleton is then segmented into regions of interest, according to the JHU atlas, and voxel measures are averaged across these regions, leaving us with a series of regional metrics (25 in total, averaging across left and right hemispheres) that can be compared between AN and HC groups. Detailed information about the pipeline and the different steps is available here: <https://github.com/pnlbwh/TBSS/blob/master/docs/TUTORIAL.md>)

### Python versions

You will need python version 3 to operate the TBSS pipeline, which can be downloaded from here: https://www.python.org/downloads/

Whether or not you already have python3, it is not necessarily the default python program. You can test which python is the default by opening a terminal and typing python --version, which will in turn print the default version of python installed (here 3.9.13):



If a python3 installation is not the default, find where it is by typing the following into the terminal:

which python3

This will return the path to your python3 installation (e.g., /usr/local/bin/python3)

You can then edit your python path using vi. Type the following in terminal to edit your bash\_profile:

vi ~/.bash\_profile

next enter the following line in the file that opens:

alias python=$path\_to\_python3installation

(In our example, $path\_to\_python3installation is /usr/local/bin/python3, so the line would be: alias python=/usr/local/bin/python3)

Quit the editor by typing: :q

You can double check this worked by typing: python --version

### Installing TBSS

Install the TBSS program by navigating to this page <https://github.com/CaitlinLloyd/TBSS_ENIGMA_AN>, and downloading the folder. Click on the code icon, and select the download zip file as before.

Once the folder is downloaded, navigate to the directory, which is called TBSS\_ENIGMA\_AN (cd $PATH\_TO\_TBSS\_DIRECTORY), and type ./install.sh setup test.

Follow instructions in the printed output from the command.

We need one extra file: the peripheral mask, which can be downloaded here: https://github.com/CaitlinLloyd/ENIGMA\_AN\_DTI/blob/main/ENIGMA\_DTI\_FA\_skeleton\_mask\_periphery.nii.gz. Once the TBSS program is installed, navigate to the enigmaDTI directory (tbss/data/enigmaDTI), and save the peripheral mask there.

### Running the TBSS pipeline

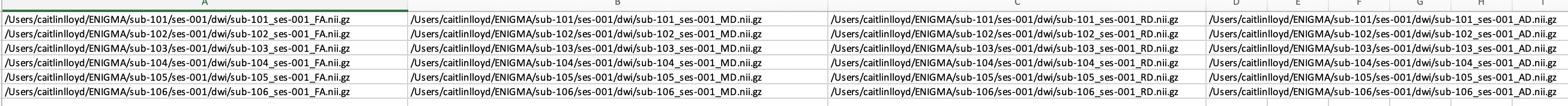
First, make sure all of your files are organized in a BIDS compliant way, with the outputs of dtifit in the dwi directory. The file structure should appear as below.

Background pattern, table

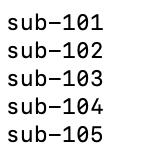
Description automatically generated

### Making the image list and subject list files

Create an IMAGELIST.csv file in excel. This file has one column per modality (e.g., FA, MD, RD, AD), and provides the full path to the file for each subject. The columns do not need headers and the file should look like this:



Then create a list of your participants and save it within a txt file (one participant per row). Here is an example:



Make sure that the participant rows in IMAGELIST.csv correspond to those in CASELIST.txt. Now you are ready to run the pipeline!

Open a terminal window, and navigate again to the tbss directory (cd $PATH\_TO\_TBSS\_DIRECTORY

Then run:

tbss/lib/tbss\_all -i $pathtoENIGMAdir/IMAGELIST.csv -c $pathtoENIGMAdir/CASELIST.txt --modality FA,MD,AD,RD --enigma -o $ENIGMAdir/$projectname/tbss\_output –qc

### QC

The addition of the qc flag allows you to manually check the registration to the ENIGMA template and remove any poorly registered images. You can also check the similarity.csv file that is found in the output directory (within the tbss directory that is now in the project directory). The values indicate the mutual information score that describes the similarity between the warped image and the ENIGMA target. Lower (i.e., more negative) scores are better. After you have completed the QC stage, click enter to continue with the pipeline.

### Bad registration (skip this section if registration is fine)

You may need to perform re-registration for some participants, or update the ENIGMA templates to your own study specific versions.

You could update the templates using various methods, and this is just one suggested method which makes a mask based on where most participants have brain tissue that overlaps with the ENIGMA templates.

Go to the FA directory created by the TBSS pipeline (cd $PATH\_TO\_PROJECT\_DIRECTORY/tbss\_output/FA/warped). Identify the images that do **not** have a good registration, and move them into a bad\_reg folder.

Create an FA mask from all of the good images.

First run fslmerge -t FA\_QC \*\_to\_target\*

And then fslmaths FA\_QC.nii.gz -bin -Tmean -thr 0.9 mean\_FA\_mask

Make a new folder for your own templates in the TBSS folder.

mkdir $PATH\_TO\_TBSS\_DIRECTORY/data/my\_templates

Copy mean\_FA\_mask.nii.gz to the my\_templates directory, Then move into this directory (cd $PATH\_TO\_PROJECT\_DIRECTORY/tbss\_output /data/my\_templates), and run the following commands:

1. fslmaths ../enigmaDTI/ENIGMA\_DTI\_FA.nii.gz -mas mean\_FA\_mask.nii.gz mean\_FA.nii.gz
2. fslmaths ../enigmaDTI/ENIGMA\_DTI\_FA\_skeleton.nii.gz -mas mean\_FA\_mask.nii.gz mean\_FA\_skeleton.nii.gz
3. tbss\_4\_prestats -0.049

You should have the following images in your folder:

mean\_DTI\_FA.nii.gz

mean\_FA\_mask.nii.gz

mean\_FA\_skeleton.nii.gz

mean\_FA\_skeleton\_mask\_dst.nii.gz

Finally, rerun the pipeline, directing it towards your new templates. You will need to tell the pipeline to use ENIGMA space and the correct lookup tables:

tbss/lib/tbss\_all -i $pathtoENIGMAdir/IMAGELIST.csv -c $pathtoENIGMAdir/CASELIST.txt --modality FA,MD,AD,RD –qc --template tbss/data/my\_temp/mean\_FA.nii.gz --skeleton tbss/data/my\_temp/mean\_FA\_skeleton.nii.gz -o $ENIGMAdir/$projectname/tbss\_output2 --labelMap JHU-ICBM-labels-1mm.nii.gz --lut tbss/data/ENIGMA\_look\_up\_table.txt --space tbss/data/enigmaDTI/

We suggest you use a new output directory (i.e., tbss\_output2) as the old one will be overwritten.

## Analysis

Now we have our output csv files describing average FA, MD, AD, RD in each of the ENIGMA ROIs as well as voxel-wise metrics. All of these files are in the stats directory of the output folder (tbss\_output) created by the TBSS pipeline. For the analysis, we use the summary files along with a file specifying covariate information.

### Making the covariates file

The Covariates.csv file should contain the full set of variables we will be controlling for or investigating in the models. A template is provided in the folder you downloaded from Github. Please fill this in using your own data. Note that missing values for Dx, Age, BMI and site are not allowed. Missing values in the other covariates are allowed, but make sure you code them as “NA” without the quotes. **Please exclude all male samples**.

* **SubjID** the IDs in the SubjID column should match the format of the IDs in the ID columns in the ROI measure files, which are also the names of the directories where the dwi images are contained (e.g., sub-101, when following BIDS formatting).
* **dx** should be a column for diagnosis (where patients are coded as 1 and controls are coded as 0). More details on which patients to include can be found [here](https://docs.google.com/document/d/16ngTjNUB_l61eJTZ64wHpI9qhfTJjacHmy8PLbB164k/edit) under the header “Analysis 1”.
* **dx3** should be a column for diagnosis (where underweight “acute” patients are coded as 2, partially weight-restored patients as 1 and controls are coded as 0. More information on how to classify your cases into the two patient categories can be found [here](https://docs.google.com/document/d/16ngTjNUB_l61eJTZ64wHpI9qhfTJjacHmy8PLbB164k/edit) under the header “Analysis 2”.
* **age** is in years at time of scan
* **bmi** is body mass index at the time of scanning.
* The **site** column details the number of sites/scanners the data were collected from. **This data will be dummy coded by the R script, so please just state the name of the site (or scanner). If there is only one site/scanner then write the name of that site for all participants.** This variable will allow you to regress out effects of scanner/site if applicable.

The following columns are optional, but will improve the quality of the results. If some variables are not applicable, i.e. you don’t have this info in your study, **please keep the variables in your Covariates.csv file** anyway but mark the values for all participants as “NA” (without quotation marks).

* **bmi\_sds** is age-adjusted BMI z\_score at time of scanning (note, you can calculate this using the R zscorer package providing you have date of birth information: <https://cran.r-project.org/web/packages/zscorer/readme/README.html>). This is not necessary for individuals who are aged 20 or older.
* **ap** refers to current antipsychotic use. Code: untreated=0, treated=1.
* **ad** refers to current antidepressants use. Code: untreated=0, treated=1.
* **ao** is age of onset in years (only needs to be filled out for the AN patients). Please include in a READ\_ME.txt file how you defined age of onset.
* **durill** is duration of illness (in years), can also be estimated as Age - AO  (only needs to be filled out for the AN patients, and please include in the READ\_ME.txt exactly how this was derived).
* **in\_out** refers to whether a patient is an in- or outpatients. Code: outpatient=0, inpatient=1.
* **comorb** refers to whether a patient has comorbid depression (=1), anxiety (=2), OCD (=3), other or a combination of comorbid psychiatric disorders (=4) or no comorbidities (=0). Please include in the READ\_ME.txt file how the diagnoses were made.
* **deprsymp** refers to a measure of depressive symptoms (Please include in the READ\_ME.txt file which measure you used)
* **anxsymp** refers to a measure of anxiety symptoms (Please include in the READ\_ME.txt file which measure you used)
* **hand** is handedness (0=Right, 1=Left, 2=Ambidextrous, NA=Not Available)
* **parentses** (highest level of parental socioeconomic status, for example based on education, occupation, or both). Please include in the READ\_ME.txt file how you defined parentses.
* **iq** is intelligence quotient. Please include in the READ\_ME.txt file how you measured IQ.
* **subtype** for AN restrictive (0) versus binge/purge subtype (1), if unknown or healthy controls (NA)

Covariates.csv should be saved to the project directory within your ENIGMA directory. The file structure would look like this:

Background pattern

Description automatically generated

Next, we run the R script to merge the ROI metrics and covariate information, and calculate the effect size for 1) differences between individuals with AN and HC, and 2) correlations between ROI metrics and clinical variables among participants with AN.

### Running the R script

The commands below assume that you are calling R from within the project folder that is within the ENIGMA folder.

You will need to direct the script to your ENIGMA and projects folders. Do this in lines 9-10 of the script:

Enigmadir in our example would be “/Users/caitlinlloyd/ENIGMA”

Projectdir is within the ENIGMA directory, so in the example here is would be “/Users/caitlinlloyd/ENIGMA/NYSPI”

Make sure the attached AN\_DTI\_betweengroup.R file is saved in the ENIGMA directory (it should be by default if you downloaded the repository from github).

On Mac/Linux you can cd to the directory on the command-line and then call R there.

cd Users/caitlinlloyd/ENIGMA/NYSPI

R

You can run the script by typing the following command (in R):

*source("Main*[*Analysis.R*](https://drive.google.com/drive/folders/1dhf3srm9qg2vZoVUboImyK6DMXLHdycW)*")*

It will print out progress to the screen, and create three big CSV files.

The CSV files store effect sizes and model statistics. The model statistics can be used for follow-up analyses (e.g, to determine what drives possibly significant interactions).

At the end of the script, the code will complete a voxel-wise analysis, in which the differences between groups at each voxel are tested. These outputs will also be saved to the project folder and can be used to determine whether there are differences at particular regions of the white matter skeleton when combining data across all sites. This step does take a while, so do not worry if it seems rather long. Also, if you interrupt this stage, you should still have the results of the main ROI-based analyses.

When you are done, exit R with q(). Then tar up the Results folder that has been created in the Project directory, and upload the zip file (named with your site) to the site results folder on the ENIGMA AN DTI [google drive](https://drive.google.com/drive/folders/1yFZXgXJKT_wg-gwiVgiNZExoJpV9_Lzu?usp=share_link).

### Local results

If you want to see the outcomes of your own analyses visually, use the script “effect\_size\_viewer.R”. This will generate plots of effect sizes.