

Paper 151-30

## A SAS® Macro for Visually Displaying Correlations between Brain Regions Using fMRI Data

Colleen M. Brensinger (1), Maximilian Herlim (1), Warren B. Bilker (1,2), Ruben C. Gur (2)

University of Pennsylvania, Philadelphia, PA

(1) Center for Clinical Epidemiology and Biostatistics and Department of Biostatistics

(2) Department of Psychiatry, Neuropsychiatry Section

### ABSTRACT

We developed a SAS macro for generating a summary graphic in 2-dimensional space of the correlations among brain regions using measures calculated from functional magnetic resonance imaging (fMRI) data. The graphical output allows the user to easily visualize differences between patient groups with regards to connectivity of brain region activity. The macro takes the output from PROC CORR and creates an annotate dataset to be used by PROC GANNO for creating the graph. Different line widths are used for connecting brain regions on the graph based on the strength of the Pearson or Spearman correlation coefficient (or, ultimately, any other measure of monotonicity). The design of this macro will be presented, along with sample output graphics.

### INTRODUCTION

In fMRI studies, procedures are used to measure hemodynamic effects of the metabolic changes that take place in an active part of the brain during a task performed by a research participant. Measurements are taken at regular intervals during the procedure. Then, voxels that show task associated changes are associated with specific regions. Various summary measures are calculated across the time series that constitutes the hemodynamic response function (HRF), such as positive area under the curve, negative area under the curve, maximum observed value, minimum observed value, etc. For these summary measures, correlations between different brain regions can then be examined in order to determine inter-regional connectivity as reflected in the linear relationships among the brain activations in the different brain regions. We were interested in 24 specific brain regions that were activated in an auditory attentional paradigm. Hence the correlation matrix generated by PROC CORR is 24x24 containing 276 correlations to be examined, for each summary measure considered for each patient subgroup. We wanted to be able to summarize these data visually. This paper presents a SAS macro for creating a summary graph of the correlations among 24 selected brain regions of interest. Output from PROC CORR is used to create an annotate dataset for input to PROC GANNO for creating the graph.

### MACRO DESCRIPTION

Prior to calling the macro, the user has the calculated summary measures for each brain region stored in a SAS dataset, with the suffix of the variable names indicating the brain region, and the prefix indicating the summary measure calculated. For example, the variable name for the negative area under the curve for the left thalamus is "auc\_n\_ltha". Table 1 describes the user inputs into the macro.

**TABLE 1:** Macro Input Parameters

Macro Variable	Description
indata	Name of the permanent SAS dataset with the brain region data to be correlated (required).
inpath	The directory containing the input dataset (required).
keep	List of the regions that you want to keep in the plot (optional). The default is to keep all 24 regions. However, if the input dataset contains only a subset of the regions, just this subset will get plotted.
drop	List of the regions that you want to drop in the plot (optional). The default is to keep all 24 regions. However, if the input dataset contains only a subset of the regions, just this subset will get plotted.
prefix	The prefix for the variable names (required). For example, auc_n_, auc_p_, maxa_
corrtype	P for Pearson or S for Spearman correlation coefficients (optional). The default is P.
graphpath	The directory in which you would like the output graphics file to be stored (required).
graphtype	The type of output graphics file desired (optional). Options are PDF, JPG, and TIFF. The default is PDF.
cut1	The lower cutpoint for determining the line thickness to be drawn (optional). The default is 0.50.

Macro Variable	Description
cut2	The mid cutpoint for determining the line thickness to be drawn (optional). The default is 0.65.
cut3	The upper cutpoint for determining the line thickness to be drawn (optional). The default is 0.80.
graphname	The name of the output graphics file (required).
graphtitle	The title to use on the graph (optional).

The outline of the macro is as follows:

- 1) Set up the 2-dimensional grid for plotting the 24 brain regions.
- 2) Create a list of the variables of interest, starting with all variables in the "indata=" input dataset, and using the "drop=", "keep=", and "prefix=" inputs to the macro to subset the summary measure of interest and the brain regions of interest.
- 3) Run PROC CORR on the variable list to output the matrix of Pearson or Spearman correlation coefficients.
- 4) Use this output dataset to create an annotate dataset.
- 5) Use PROC GANNO to create the graph specified in the annotate dataset.

The macro starts by setting up a 2-dimensional grid for the 24 brain regions of interest, along with the abbreviations that will be used for labeling the regions on the plot. Table 2 lists the brain regions and the coordinates on the x- and y-axis that we chose. All regions in the left hemisphere have negative x-coordinates, while the right hemisphere has positive x-coordinates.

**TABLE 2:** Brain Regions with X- and Y- axis coordinates

Brain Region	Variable suffix	Abbreviation on graph	X-coordinate	Y-coordinate
Left Frontal Medial	LMF	MF	-3	43
Right Frontal Medial	RMF	MF	3	43
Left Thalamus	LTHA	THA	-3	5
Right Thalamus	RTHA	THA	3	5
Left Insula	LINS	INS	-7	7
Right Insula	RINS	INS	7	7
Left Cingulate	LCG	CG	-13	11
Right Cingulate	RCG	CG	13	11
Left Parietal Postcentral	LPAPOSTC	PaPostC	-11	31
Right Parietal Postcentral	RPAPOSTC	PaPostC	11	31
Left Temporal Lobe, Transverse Gyrus	LTETRA	TeTra	-17	40
Right Temporal Lobe, Transverse Gyrus	RTETRA	TeTra	17	40
Left Inferior Parietal	LIP	IP	-19	26
Right Inferior Parietal	RIP	IP	19	26
Left Parietal Precuneus	LPAPRECU	PaPrecu	-21	23
Right Parietal Precuneus	RPAPRECU	PaPrecu	21	23
Left Occipital Lingual	LOCLIN	OClin	-16	14
Left Occipital Middle	LOCMID	OCmid	-23	16
Left Occipital Cuneus	LOCCU	OCcu	-22	18
Left Parahippocampal Gyrus	LPHG	PHG	-26	28
Left Middle Temporal Gyrus	LMTG	MTG	-25	33
Left Superior Temporal Gyrus	LSTG	STG	-11	35
Left INF Frontal	LIF	IF	-9	38
Left Superior Frontal	LSF	SF	-9	45

Next, the macro uses the OUT= option of PROC CONTENTS to get a list of the variable names in the input dataset. It uses the user specified prefix to keep only the summary measures of interest, ignoring any of the other variables in the input dataset. For example, if prefix=auc\_n, then any variables beginning with, say, "maxa\_" will be ignored.

Occasionally, the user may want to exclude some of the 24 brain regions from the graph. For instance, if fMRI data taken during visual tasks is being examined, a different subset of brain regions may be of interest compared to when the fMRI data is taken during auditory tasks. The keep= or drop= options can accommodate this, allowing the user to specify an entire region to be dropped or just the left hemisphere of a particular region (or vice versa). If drop=rtha, just the right thalamus will be excluded from the graph, while the left thalamus will be plotted. On the other hand, if drop=tha, both the

right and left thalamus regions will be excluded. Similarly with the keep= option. If keep=tha, both left and right thalamus will be used, but if keep=ltha, only the right thalamus will be excluded. The macro uses the index function to check each of the variable names in the input dataset against the drop= and keep= macro variables to determine which variables to retain in the analysis.

These variables of interest get saved to a macro variable called “corrvars”, which gets passed into PROC CORR. Using either the OUTP= or OUTS= option of PROC CORR, the correlation matrix is saved to a SAS dataset. The correlations for each of the pairwise combinations of brain regions are examined. Any weak correlations that are less than the cut1= cutpoint in absolute value are not connected on the graph. The remaining are used to draw connecting lines between the brain regions which are strongly associated with each other, using different line widths based on the cut1, cut2, and cut3 user inputs. Positive correlations are connected using solid lines, while negative correlations are connected with dashed lines. The annotate dataset contains the x- and y- coordinates for each brain region and the label for each brain region, plus instructions for connecting any of the brain regions which are strongly correlated. It also contains the graph title specified in the graphtitle= option, and a legend for the cutpoints used for determining the different line widths. Finally, the annotate dataset is passed into PROC GANNO, which creates the final graph. The graph is stored to an output graphics file specified by the graphname= and graphpath= input options. The graph can be stored in PDF, JPG, or TIFF format based on the input option graphtype=.

### EXAMPLE

The fMRI data was collected on 32 patients with schizophrenia and 46 healthy controls seen at the Brain Behavior Lab at the University of Pennsylvania. The data was measured while participants were performing an auditory attentional task (standard P300). The maximum activation in each brain region was calculated and the goal is to see which of the 24 brain regions are correlated with each other. This produces a 24x24 symmetric correlation matrix for the patients with schizophrenia, as displayed in Figure 1. It is very tedious to examine each of the 276 pairwise correlations to determine which brain regions are correlated with each other, and to what degree. The following macro call was used to create the summary graph in Figure 2 for the schizophrenic patients, and a separate, but similar call was used to generate the graph for the healthy controls.

```
%brainconnect (indata=schizpats,
               inpath=/project/schiz/rubengur/sas_graph/brain_connect,
               keep=,
               drop=,
               prefix=maxa_,
               corrtype=P,
               graphpath=/project/schiz/rubengur/sas_graph/brain_connect,
               graphtype=pdf,
               cut1=0.50,
               cut2=0.65,
               cut3=0.80,
               graphname=schizophrenics,
               graphtitle=Maximum: Schizophrenics);
```

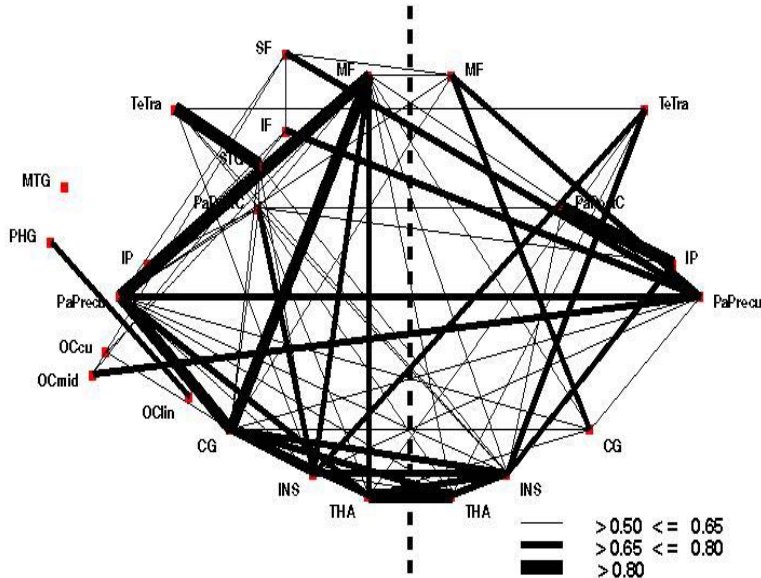
From the graph, we can easily see which regions most strongly correlated with each other in patients with schizophrenia (e.g. Left Frontal Medial with Left Parietal Precuneus, Left Frontal Medial with Left Cingulate, Left Parietal Precuneus with Left Cingulate, Left Superior Temporal Gyrus with Left Temporal Transverse gyrus (TeTra), Right Inferior Parietal with Right Parietal Postcentral, Left thalamus with Right Thalamus), and in healthy people (Left thalamus with Right Thalamus, Left Parahippocampal Gyrus with Left Occipital Lingual). It also becomes evident that for healthy controls, there are far fewer correlations among brain regions. The activation in patients seems to indicate “hyperconnected” brains.

**FIGURE 1:**

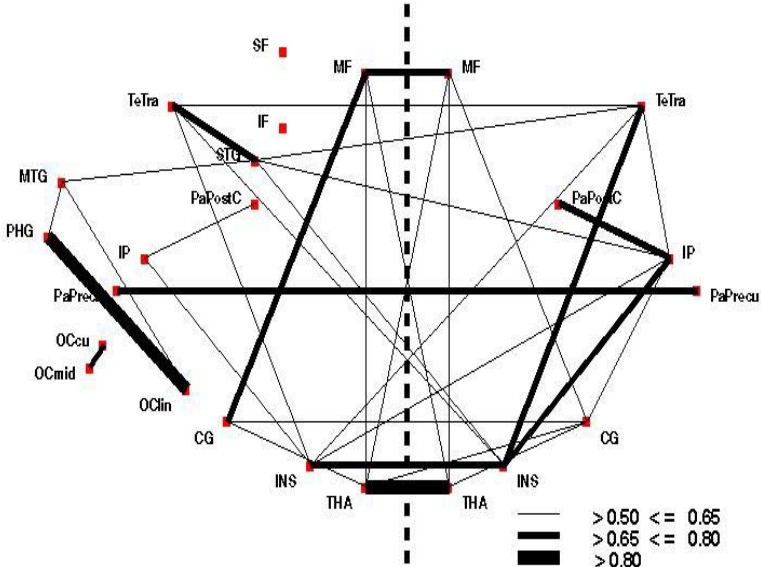
	maxa_LIF	maxa_LIP	maxa_LMF	maxa_LPaPostC	maxa_LPaPrecu	maxa_LSTG	maxa_RIP	maxa_RMF	maxa_RPaPostC	maxa_RPaPrecu	maxa_LCG	maxa_LINS	maxa_LMTG	maxa_LOCCu	maxa_LOClin	maxa_LOCmid	maxa_LPHG	maxa_LSF	maxa_LTHA	maxa_lTeTra	maxa_rCG	maxa_rINS	maxa_rTHA	maxa_rTeTra
maxa_LIF	1.00	0.16	0.43	0.22	0.62	0.35	0.07	0.44	-0.08	0.74	0.39	0.44	0.14	0.41	0.43	0.57	0.38	0.61	0.22	0.22	0.27	0.13	0.30	0.18
maxa_LIP	0.16	1.00	0.10	0.61	0.13	0.52	0.37	0.11	0.32	-0.05	0.15	0.38	-0.24	-0.04	-0.18	-0.22	-0.23	-0.13	0.16	0.40	0.33	0.12	0.16	0.19
maxa_LMF	0.43	0.10	1.00	0.58	0.84	0.41	0.47	0.59	0.35	0.59	0.95	0.70	0.08	0.54	0.23	0.33	0.25	0.39	0.65	0.37	0.52	0.61	0.62	0.30
maxa_LPaPostC	0.22	0.61	0.58	1.00	0.50	0.65	0.56	0.26	0.53	0.22	0.63	0.68	-0.10	0.26	0.11	0.03	0.05	-0.04	0.53	0.64	0.49	0.57	0.58	0.41
maxa_LPaPrecu	0.62	0.13	0.84	0.50	1.00	0.48	0.32	0.62	0.19	0.79	0.82	0.64	0.04	0.48	0.28	0.37	0.35	0.52	0.67	0.46	0.51	0.51	0.64	0.33
maxa_LSTG	0.35	0.52	0.41	0.65	0.48	1.00	0.26	0.10	0.26	0.29	0.52	0.62	-0.04	0.30	0.26	-0.00	0.04	0.03	0.42	0.87	0.24	0.56	0.40	0.48
maxa_RIP	0.07	0.37	0.47	0.56	0.32	0.26	1.00	0.19	0.87	0.06	0.49	0.43	-0.07	0.22	-0.19	-0.07	-0.11	-0.20	0.47	0.24	0.24	0.66	0.47	0.36
maxa_RMF	0.44	0.11	0.59	0.26	0.62	0.10	0.19	1.00	0.14	0.67	0.51	0.43	-0.18	0.23	0.08	0.35	0.26	0.57	0.48	0.10	0.80	0.23	0.46	0.30
maxa_RPaPostC	-0.08	0.32	0.35	0.53	0.19	0.26	0.87	0.14	1.00	-0.08	0.42	0.41	-0.16	0.03	-0.11	-0.14	-0.08	-0.34	0.43	0.36	0.31	0.62	0.41	0.40
maxa_RPaPrecu	0.74	-0.05	0.59	0.22	0.79	0.29	0.06	0.67	-0.08	1.00	0.52	0.53	0.09	0.38	0.38	0.69	0.47	0.80	0.47	0.24	0.53	0.29	0.43	0.36
maxa_LCG	0.39	0.15	0.95	0.63	0.82	0.52	0.49	0.51	0.42	0.52	1.00	0.73	0.07	0.55	0.19	0.29	0.18	0.34	0.73	0.48	0.53	0.69	0.69	0.40
maxa_LINS	0.44	0.38	0.70	0.68	0.64	0.62	0.43	0.43	0.41	0.53	0.73	1.00	-0.10	0.43	0.23	0.26	0.04	0.28	0.61	0.58	0.47	0.70	0.56	0.68
maxa_LMTG	0.14	-0.24	0.08	-0.10	0.04	-0.04	-0.07	-0.18	-0.16	0.09	0.07	-0.10	1.00	-0.04	0.01	0.26	0.04	0.18	-0.10	-0.11	-0.05	-0.17	-0.16	-0.28
maxa_LOCCu	0.41	-0.04	0.54	0.26	0.48	0.30	0.22	0.23	0.03	0.38	0.55	0.43	-0.04	1.00	0.02	0.43	0.22	0.32	0.24	0.16	0.06	0.43	0.35	0.07
maxa_LOClin	0.43	-0.18	0.23	0.11	0.28	0.26	-0.19	0.08	-0.11	0.38	0.19	0.23	0.01	0.02	1.00	0.22	0.65	0.25	0.22	0.38	-0.00	0.18	0.20	0.27
maxa_LOCmid	0.57	-0.22	0.33	0.03	0.37	-0.00	-0.07	0.35	-0.14	0.69	0.29	0.26	0.26	0.43	0.22	1.00	0.45	0.62	0.17	-0.16	0.32	0.06	0.15	0.08
maxa_LPHG	0.38	-0.23	0.25	0.05	0.35	0.04	-0.11	0.26	-0.08	0.47	0.18	0.04	0.04	0.22	0.65	0.45	1.00	0.39	0.35	0.14	0.16	0.07	0.35	-0.06
maxa_LSF	0.61	-0.13	0.39	-0.04	0.52	0.03	-0.20	0.57	-0.34	0.80	0.34	0.28	0.18	0.32	0.25	0.62	0.39	1.00	0.26	-0.01	0.47	-0.04	0.22	0.12
maxa_LTHA	0.22	0.16	0.65	0.53	0.67	0.42	0.47	0.48	0.43	0.47	0.73	0.61	-0.10	0.24	0.22	0.17	0.35	0.26	1.00	0.42	0.52	0.67	0.91	0.59
maxa_lTeTra	0.22	0.40	0.37	0.64	0.46	0.87	0.24	0.10	0.36	0.24	0.48	0.58	-0.11	0.16	0.38	-0.16	0.14	-0.01	0.42	1.00	0.27	0.56	0.42	0.50
maxa_rCG	0.27	0.33	0.52	0.49	0.51	0.24	0.24	0.80	0.31	0.53	0.53	0.47	-0.05	0.06	-0.00	0.32	0.16	0.47	0.52	0.27	1.00	0.22	0.42	0.38
maxa_rINS	0.13	0.12	0.61	0.57	0.51	0.56	0.66	0.23	0.62	0.29	0.69	0.70	-0.17	0.43	0.18	0.06	0.07	-0.04	0.67	0.56	0.22	1.00	0.68	0.66
maxa_rTHA	0.30	0.16	0.62	0.58	0.64	0.40	0.47	0.46	0.41	0.43	0.69	0.56	-0.16	0.35	0.20	0.15	0.35	0.22	0.91	0.42	0.42	0.68	1.00	0.51
maxa_rTeTra	0.18	0.19	0.30	0.41	0.33	0.48	0.36	0.30	0.40	0.36	0.40	0.68	-0.28	0.07	0.27	0.08	-0.06	0.12	0.59	0.50	0.38	0.66	0.51	1.00

FIGURE 2:

Maximum value: Schizophrenics



Maximum value: Controls



**CONCLUSION**

The graphical display of the correlations among brain regions created by this macro eases the interpretation of complex data. Separate graphs for subgroups (e.g. patients vs. healthy controls) can be easily created, which can visually highlight differences between groups. Future efforts will explore generating 3-dimensional graphs of the brain regions based on standard atlas coordinates.

The SAS code is available upon request from the authors.

**REFERENCES**

SAS Institute Inc., SAS Macro Language: Reference, First Edition, Cary, NC: SAS Institute Inc., 1997.  
SAS Institute Inc., SAS/GRAPH® Software: Reference, Version 8, Cary, NC: SAS Institute Inc., 1999.

**CONTACT INFORMATION**

The authors welcome feedback via e-mail at [cbrensin@cceb.upenn.edu](mailto:cbrensin@cceb.upenn.edu)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.