Check for updates

METHOD ARTICLE

Detecting variable responses within fMRI time-series of volumes-of-interest using repeated measures ANOVA [version 1; peer review: 2 approved with reservations]

Paul M. Macey^{1,2}, Philip J. Schluter³, Katherine E. Macey¹, Ronald M. Harper⁴

¹UCLA School of Nursing, LA, CA, USA

²Brain Research Institute, Department of Neurobiology, David Geffen School of Medicine, UCLA, LA, CA, USA

³School of Health Sciences, University of Canterbury, Christchurch, New Zealand

⁴Department of Neurobiology, David Geffen School of Medicine, UCLA, LA, CA, USA

V1 First published: 05 Apr 2016, 5:563
<https://doi.org/10.12688/f1000research.8252.1>

Latest published: 08 Jul 2016, 5:563
<https://doi.org/10.12688/f1000research.8252.2>

Abstract

We present an approach to analyzing fMRI timetrends from volumes-of-interest (VOI) within and between subject groups using repeated measures analysis of variance (RMANOVA), which allows temporal patterns to be examined without an *a priori* model of expected timing or pattern of response. The method serves as a complement to whole-brain voxel-based analyses, and is useful for detecting complex responses within pre-determined brain regions, or as a post-hoc analysis of regions of interest identified by whole-brain assessments. We illustrate an implementation of the technique in the statistical software package SAS. VOI timetrends are extracted from conventionally preprocessed fMRI images. A timetrend of average signal intensity across the VOI during the scanning period is calculated for each subject. The values are scaled relative to baseline periods, imported into SAS, and the procedure PROC MIXED implements the RMANOVA. The ensuing results allow determination of significant overall effects, and time-point specific within- and between-group responses relative to baseline. We illustrate the technique using fMRI data from two groups of subjects who underwent a respiratory challenge. RMANOVA allows insight into the timing of responses and response differences between groups, and so is suited to fMRI paradigms eliciting complex response patterns.

Keywords

mixed effect models , regression analysis , statistical models , physiological responses , functional magnetic resonance imaging

Open Peer Review

Approval Status

	1	2
version 2		
(revision)
08 Jul 2016
version 1
05 Apr 2016

1. **Joke Durnez**, Stanford University, Stanford, USA

2. **Matthew Brett**, University of California, Berkeley, Berkeley, USA

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **INCF** gateway.

Corresponding author: Paul M. Macey (pmacey@ucla.edu)

Competing interests: No competing interests were disclosed.

Grant information: Supported by the National Institutes of Health National Institute of Nursing Research NR 013693.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2016 Macey PM *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Macey PM, Schluter PJ, Macey KE and Harper RM. **Detecting variable responses within fMRI time-series of volumes-of-interest using repeated measures ANOVA [version 1; peer review: 2 approved with reservations]** F1000Research 2016, 5:563 <https://doi.org/10.12688/f1000research.8252.1>

First published: 05 Apr 2016, 5:563 <https://doi.org/10.12688/f1000research.8252.1>

Introduction

We describe a procedure for analyzing fMRI time-courses across multiple subjects in pre-determined regions without an *a priori* pattern or timing of expected response. Functional MRI analyses usually require defining a model and identifying voxels whose time courses follow that model; other model-free event-related approaches test for responses that are coincident in time or duration. However, certain tasks elicit neural responses that vary in pattern and timing. For example, responses to physiological challenges typically occur across neural structures in a sequence rather than simultaneously, and the patterns differ according to structure¹⁻³. While existing software like SPM does allow for model-free assessment of responses, such as identifying the hemodynamic response based on a finite impulse response function in SPM12⁴, those approaches can be complex to implement and interpret. The present method arose out of our research group's need to perform basic time-trend analyses of fMRI signals in specific regions.

The approach we propose is analyzing fMRI time series using repeated measures analysis of variance (RMANOVA), allowing for detection of within-group responses relative to a baseline or resting state, and of between-group differences in response. As with any analysis of variance (ANOVA), RMANOVA tests the equality of means that are assumed to be normally distributed. However, RMANOVA also accounts for the correlation between repeated measures within subjects, whereas ANOVA does not. In addition to assessment of significant reactions relative to a baseline period, or of significant group differences in response, the method allows for identifying specific time-points when these differences occur. Therefore, transient, late-developing, and complex sequences of increasing and decreasing signal changes can be distinguished with RMANOVA. The technique is suited to analysis of longer task paradigms evoking a complex pattern of responses. We present an implementation of RMANOVA in the programming language SAS (SAS Institute Inc., Cary, NC).

Description of procedure

Functional MRI data collected in one or more groups of subjects are preprocessed, typically with slice timing and motion correction, and the images may be spatially normalized. An additional, optional step is detrending to remove global effects. In traditional cluster analyses, global effects may be accounted for within the analysis steps using, for example, SPM⁵. Intensity normalization using proportional scaling has been the most common approach to removing such effects, but for the example presented here, removal of global signal changes is performed on the preprocessed images prior to analysis with RMANOVA using a custom technique⁶.

Volumes-of-interest (VOI) are defined, usually by outlining a structure of interest on an image volume and saving that outline as a binary image, or mask. The VOI may be defined individually on a subject-by-subject basis, or globally for all subjects on a template (the latter case assumes subjects' images are accurately spatially normalized to the template). The mask is used to extract the intensity of only those voxels within the VOI, at each time point in the fMRI series. For each time point, the average of those voxel values is recorded, resulting in a time-series corresponding to the blood oxygen level dependent (BOLD) signal time course of that VOI in

the fMRI series. The procedure is repeated for all subjects, resulting in one time trend for each subject per VOI. The time trends may optionally be deconvolved with a hemodynamic response function (HRF), allowing the timing of significant responses to be inferred as neural, rather than BOLD. The resulting VOI time trends are subjected to RMANOVA.

Repeated measures ANOVA can be implemented using a mixed effect model, a generalization of the standard linear model which allows for correlation and non-constant variability within the data⁷. The methodology described here assumes use of the SAS function PROC MIXED, although other software packages also perform RMANOVA. However, SAS is very flexible, and includes a wide range of diagnostics tests. (See white paper "Comparing the SAS® GLM and MIXED Procedures for Repeated Measurements Analysis," Russ Wolfinger and Ming Chang, SUGI Proceedings, 1995; available at <https://support.sas.com/rnd/app/stat/papers/abstracts/mixedglm.html>). The statistical procedures calculate the variance-covariance matrix based on the dependencies defined by repeated measurements within subjects, and by group classification of subjects in the case of more than one group (REPEATED and GROUP options within PROC MIXED; by convention, SAS syntax is in upper case)⁷.

A classification variable (t) representing baseline and time-points in the task or challenge period is created, with values of t during the baseline or reference period constant, and values increasing for each time point in the series. For example, we have collected fMRI data during physiological challenges with one minute of baseline followed by 90 sec of challenge, at a repetition time of 6 seconds (25 volumes total); thus, t is 0 for the first 10 scans, and 1 for the 11th volume, 2 for the 12th, and so on until 15 for the 25th. Note that for data collected at higher sampling rates or across longer time periods, the time-classification variable may include multiple data points, as opposed to the individual data points of the fMRI time-series; for example, we have classified instantaneous (beat-by-beat) heart rates recorded throughout a two minute period into twelve 10 sec epochs⁸. If more than one group of subjects is being analyzed, a second classification variable is defined at each time point with a value representing the group to which that series belongs. An example of such a format in an Excel file is shown in Appendix A; these data may be imported into SAS.

Before assessing the fMRI time-trend results of the RMANOVA analysis (see sections below for full details), certain residual and influence diagnostic tests should be performed to ensure that the model does not seriously violate underlying assumptions. A relevant subset of such tests is described in sections below. Many practical applications of RMANOVA violate the model assumptions to some degree, but the approach is considered robust to moderate departures of these assumptions. In the present application to fMRI data, a key test is to ensure that the residual distributions are not seriously skewed. We suggest the following checks. First determine the s residuals, which are scaled so that more than 95% should fall within the critical limits of the appropriate Student's t distribution, which for most purposes can be assumed to be within ± 1.96 (and centered on 0). The shape of the histogram for each group of the studentized residuals should be bell-shaped. There should also be no effect

of time, so each group plot of the studentized residuals by time should not show a trend. We recommend using locally-weighted-scatter-plot (lowess) curves to identify such trends. In the case where the diagnostic tests fail, there may still be some value in continuing with the model, but we suggest presenting the results of the residual diagnostics along with the findings, to allow readers to assess for themselves the validity of the model. Further options are explained in sections below, along with other possible diagnostic tests.

There are two levels of results produced by RMANOVA: the significance of the overall model at the group, time, and group-by-time levels, and the significance of between-group and within-group effects at each time point (see section below). According to the Tukey-Fisher criterion for multiple comparisons, the model is first assessed at the overall level for significance before investigating time-point specific effects; i.e., if the overall group-by-time effect is not significant, the between-group effects at each time point are not considered, regardless of their reported significance level. Similarly, if the overall time effect is not significant, then within-group effects at each time point are not considered.

Specific programming details are presented in the next sections to facilitate replication.

Details of Excel format for exporting to SAS

Data are arranged in rows per subject. Each subject is identified by a unique classification variable (SUBJECT, which may be

the subject name), and similarly, the group to which that subject belongs is designated by the GROUP classification variable e.g., 1, 2, ...etc.). The fMRI times series values are arranged in adjacent columns, with one column for each time point (image volume), and with the value representing the average of the voxels within the VOI at that time point for that subject (row in [Figure 1](#)). Values are in percentage change relative to a baseline period. The column headings for the time points take the format of a root ("val" in [Figure 1](#) and the example code in next section) followed by a number indicating the sequence of values, starting at 0 (e.g., val0, val1, val2, ...etc.).

The SAS code for importing these data into a library "voilib" is as follows:

```
LIBNAME voilib '[sas_library_folder]';

PROC IMPORT OUT=voilib.data
  DATAFILE = '[fn]'
  DBMS=EXCEL2000 REPLACE;
  RANGE = '[worksheet]'
  GETNAMES=YES;

RUN;
```

where [sas_library_folder] is the path to where the SAS library files will be stored, [fn] is the Excel file name, and [worksheet] is the name of the worksheet within the Excel file containing the data.

SUBJECT	GROUP	val0	val1	val2	val3	val*
SUBG1_01	1	0.169338438	0.690160019	-0.090090898	0.035881106	...
SUBG1_02	1	0.066887927	-0.047821325	0.213851773	0.999991044	...
SUBG1_03	1	0.243283375	0.098082439	-0.047729531	-0.028972177	...
SUBG1_04	1	0.161134685	0.542903961	0.045133575	-0.04391523	...
SUBG1_05	1	0.439412099	0.074224953	-0.377516691	-0.050213873	...
SUBG1_06	1	-0.100684038	0.118186715	0.18986852	0.141918682	...
SUBG1_07	1	0.047130127	-0.041779977	-0.233438639	-0.205252945	...
SUBG1_08	1	-0.042619783	0.772760009	1.060817731	0.096007406	...
SUBG1_09	1	0.632159414	0.433151981	0.392145591	0.002595301	...
SUBG1_10	1	-0.510494348	0.221567091	0.51953683	-0.01247043	...
SUBG1_11	1	-0.640114016	-0.043465916	-0.778802406	-0.785306512	...
SUBG1_12	1	-0.100064051	0.095994832	-0.106819119	0.388850141	...
SUBG1_13	1	-0.05285291	-0.339680093	-0.038541328	0.000104044	...
SUBG1_14	1	0.027030525	-0.285822204	-0.694550393	0.420641691	...
SUBG2_01	2	-0.076203224	0.160403878	-0.34543254	0.288751386	...
SUBG2_02	2	0.676821539	-0.928531366	0.032705925	-0.939375044	...
SUBG2_03	2	-0.270182331	-1.375413081	0.605061367	-0.519048051	...
SUBG2_04	2	-0.835982459	-0.260346679	-0.382581264	-0.076481006	...
SUBG2_05	2	-0.986305852	-0.847891379	-0.790110325	-0.883754156	...
SUBG2_06	2	0.015715786	-0.172088373	-0.086682743	0.344163109	...
SUBG2_07	2	-0.090998421	-0.128192512	0.700970591	0.062223273	...
SUBG2_08	2	-0.420830922	0.615435207	0.641868118	-0.081248131	...
SUBG2_09	2	0.701484744	-0.088164329	0.346950213	-0.158094864	...
SUBG2_10	2	-0.09863948	-0.189701591	-0.434214693	-0.045334386	...
SUBG2_11	2	1.044063133	0.121255576	-0.02955588	-0.003427736	...
SUBG2_12	2	-0.421111957	-0.011023711	0.021255198	-0.129319237	...

Figure 1. Example of data in Excel spreadsheet ready for importing into SAS.

SAS can also import data from a text file, which is useful for series with more than 250 time points, or for systems without Excel. The SAS help has details on importing such files.

Note that SAS code is not case dependent. In the examples presented here, capitals are used to indicate SAS commands, and bolded names within square brackets are used to indicate user-specified variables. The remaining lowercase names are variables specific to these examples.

SAS code for RMANOVA

A simple example of formatting the data is illustrated with the following SAS code, where a variable “epoch” classifies all time points as baseline epoch or epoch number corresponding to task scan number, with the assumption that one baseline period is followed by the task period:

```
PROC FORMAT;
    VALUE ynft 1='yes' 0='no';
    VALUE gft 1='[group1]' 2 = '[group2]';
RUN;
DATA voilib.formated; SET voilib.data;
    ARRAY y{[num_tp]} val0 - val [num_tp-1];
    DO i=1 TO [num_tp];
        time=i; val=y{i};
        IF 0 < time <= [bl] THEN epoch=[bl];
        IF time > [bl] THEN epoch=time;
        IF 0 < time <= [bl] THEN chal='baseline';
        IF time > [bl] THEN chal='challenge';
    OUTPUT;
END;
KEEP subject group time epoch chal;
RUN;
```

where [num_tp] is the number of values in the time-series and [bl] is the last baseline scan. The names for each group may be entered in place of [group1] and [group2]. The above formatting code can be modified according to the fMRI task timing. For time series with large numbers of samples, the epoch variable can be used to combine multiple values within a time period.

The RMANOVA, along with formatting and output filtering, is implemented as follows:

```
PROC MIXED NOITPRINT NOCLPRINT DATA= voilib.formated
EMPIRICAL;
    CLASS group epoch subject;
    MODEL val = group epoch group*epoch / OUTP = prediction
    RESIDUAL;
    REPEATED / TYPE=CS SUB=subject(group) group=group;
    LSMEANS group*epoch / DIFF;
```

```
ODS LISTING EXCLUDE LSMEANS;
ODS OUTPUT LSMEANS= voilib.means;
ODS LISTING EXCLUDE DIFFS;
ODS OUTPUT DIFFS= voilib.differences;
ODS OUTPUT COVPARMS= voilib.covparams;
ODS OUTPUT LRT= voilib.likelihoodratio;
ODS OUTPUT TESTS3= voilib.type3tests;
```

RUN;

where the italicized statements define the model, and the remaining control the output (including residuals). Note that the RESIDUAL option, which directs studentized residuals to be calculated, is only available from SAS version 9.1. The CLASS statement determines classification variables.

The EMPIRICAL option in the proc mixed command line uses the Huber-White sandwich estimator of variance, rather than the default, as it is more conservative but considered more robust, particularly if the assumption of normality is violated (as is often the case).

The output from this step contains numerous descriptors of the model. The “Null Model Likelihood Ratio Test” gives the probability of the null model being true (column heading “Pr > ChiSq”). If the model passes the Tukey-Fisher criterion for multiple comparisons, i.e., the probability of the null model is less than 0.05, the model is investigated at the variable level using the “Type 3 Tests of Fixed Effects” output. Probabilities of the null contribution to the model by individual independent variables are presented in the “Pr > F” column for the group, epoch and group*epoch. “Group” represents a group effect, which is often not significant for fMRI data presented in percent change relative to baseline. The variable epoch represents a time effect, corresponding to within-group significant responses. The group*epoch term represents a group-by-time effect, corresponding to between-group differences across time. Again considering the Tukey-Fisher criterion, if for any of these variables the likelihood of the null model is greater than 0.05, then that variable is determined to be not significant. If less than the threshold, the model is investigated further at each time point.

Further code is needed to select the relevant output for determining the time points of within and between group significant differences. The code below is one of several possible ways to extract the relevant results.

```
DATA voilib.means; SET voilib.means;
    RENAME estimate=adjmean;
    DROP effect df tValue Probt;
RUN;
PROC SORT DATA= voilib.means; BY group epoch; RUN;
DATA voilib.differences; SET voilib.differences;
    DROP effect df;
    RENAME estimate=meandiff;
```



```

LABEL estimate='mean diff' _epoch='vs epoch'
_group='vs group' probt='p value';

```

```

RUN;

```

```

DATA voilib.betweengroup; SET voilib.differences;

```

```

IF epoch ne _epoch THEN delete;

```

```

DROP _epoch;

```

```

RUN;

```

```

PROC SORT DATA= voilib.betweengroup; BY epoch group; RUN;

```

```

DATA voilib.withingroup; SET voilib.differences;

```

```

IF group NE _group THEN delete;

```

```

DROP _group;

```

```

IF epoch NE 28 THEN delete;

```

```

RUN;

```

```

PROC SORT DATA= voilib.withingroup; BY group _epoch; RUN;

```

The above code will create Withingroup and Betweengroup tables in the SAS library voilib.

If “epoch” showed a significant effect in the Type 3 tests (above), the time points where this difference appeared may be determined by examining the Withingroup table. As shown in Figure 2, each

row in the table compares baseline (labeled “epoch”) vs subsequent epochs (labeled “vs epoch”), with a mean difference, standard error and *P* value; if the latter indicates significance (< 0.05), that time point is considered to have a significant response relative to baseline for that group.

Similarly, if “group*epoch” showed a significant effect in the Type 3 tests (above), the time-points where this difference appeared may be determined by examining the Betweengroup table. As shown in Figure 3, each row corresponds to a time point (labeled “epoch”) and compares one group vs the other, with a mean difference, standard error and *P* value; if the latter indicates significance (< 0.05), that time point is considered to have a significant group difference.

Assumptions, diagnostic checks, and limitations

The RMANOVA method has a number of underlying assumptions which should be considered. The assumptions of the RMANOVA method which differ from ANOVA are that 1) the means are linear over time, 2) multivariate normality, 3) homogeneity of covariance matrices, and 4) independence. The method is reasonably robust to violations of the second and third assumptions. Violations of independence produce a non-normal distribution of the residuals, which invalidates the F-ratio. The most common violations of independence occur when either random selection or random assignment is not used or the compound symmetry covariance assumption is inappropriate. Violation of the homogeneity of covariance matrices generally results in the overall test having a higher Type I error rate than nominally set.

	GROUP	epoch	vs epoch	mean diff	Standard Error	t Value	p value
1	GROUP1	28	29	0.4552	0.2154	2.11	0.0351
2	GROUP1	28	30	0.002794	0.2154	0.01	0.9897
3	GROUP1	28	31	0.7125	0.2154	3.31	0.0010
4	GROUP1	28	32	0.4379	0.2154	2.03	0.0426
5	GROUP1	28	33	0.5178	0.2154	2.40	0.0166
6	GROUP1	28	34	0.09533	0.2154	0.44	0.6583
7	GROUP1	28	35	0.7120	0.2154	3.31	0.0010
8	GROUP1	28	36	-0.03375	0.2154	-0.16	0.8756
9	GROUP1	28	37	0.04744	0.2154	0.22	0.8258
10	GROUP1	28	38	0.1753	0.2154	0.81	0.4163
11	GROUP1	28	39	0.4913	0.2154	2.28	0.0230
12	GROUP1	28	40	0.2444	0.2154	1.13	0.2572

Figure 2. Example of within-group SAS table of time-points of difference relative to baseline.

The screenshot shows the SAS software interface. On the left, the 'Explorer' window displays the 'Contents of 'Voilib'' with icons for 'Between...', 'Covparams', 'Data', 'Differences', 'Formatted', 'Likelihood...', 'Means', 'Type3tests', and 'Withingr...'. On the right, the 'VIEWTABLE: Voilib.Betweengroup' window displays a table with 12 rows and 8 columns. The columns are: an unlabeled index, GROUP, epoch, vs group, mean diff, Standard Error, t Value, and p value. The data shows comparisons between GROUP1 and GROUP2 at epochs 28 through 39.

	GROUP	epoch	vs group	mean diff	Standard Error	t Value	p value
1	GROUP1	28	GROUP2	1.05E-14	0.1231	0.00	1.0000
2	GROUP1	29	GROUP2	-0.1044	0.3320	-0.31	0.7532
3	GROUP1	30	GROUP2	-0.1386	0.3320	-0.42	0.6766
4	GROUP1	31	GROUP2	-0.7244	0.3320	-2.18	0.0296
5	GROUP1	32	GROUP2	-0.08134	0.3320	-0.25	0.8065
6	GROUP1	33	GROUP2	-0.3957	0.3320	-1.19	0.2339
7	GROUP1	34	GROUP2	-0.1173	0.3320	-0.35	0.7241
8	GROUP1	35	GROUP2	-0.5478	0.3320	-1.65	0.0996
9	GROUP1	36	GROUP2	0.3822	0.3320	1.15	0.2501
10	GROUP1	37	GROUP2	0.03567	0.3320	0.11	0.9145
11	GROUP1	38	GROUP2	0.3332	0.3320	1.00	0.3160
12	GROUP1	39	GROUP2	-0.7307	0.3320	-2.20	0.0282

Figure 3. Example of between-group SAS table of time-points of difference between groups.

In the current application, a subset of key diagnostic and residual checks is suggested with the implementation of this technique.

- Plots of predicted and observed data, to visually determine any major bias.
- Residual checks of normality, including 95% of Studentized residuals falling within ± 1.96 , showing approximately normal group distributions, and scatter plots over time with lowess curves superimposed to ensure lack of group trends with time. Formal tests of normality may also be performed, although these will typically fail.

Assuming the “RESIDUAL” option in PROC MIXED has been used, which implements the studentized residuals, the following SAS code implements the tests:

```
PROC SORT DATA=voilib.formatted; BY group epoch; RUN;
```

```
PROC SORT DATA=voilib.means; BY group epoch; RUN;
```

```
PROC SORT DATA=voilib.prediction; BY group epoch; RUN;
```

```
DATA voilib.residuals;
```

```
    MERGE voilib.formatted voilib.means voilib.prediction;
```

```
    BY group epoch;
```

```
    resid = val - adjmean;
```

```
    LABEL resid='residual errors';
```

```
    absresid = abs(resid);
```

```
RUN;
```

```
PROC UNIVARIATE NORMAL PLOT DATA=voilib.residuals;
```

```
    VAR resid; BY group;
```

```
    TITLE 'Residual Analysis';
```

```
RUN;
```

```
SYMBOL1 I=NONE V=STAR; * C=BLACK; *;
```

```
PROC GPLOT DATA=voilib.residuals;
```

```
    PLOT resid*adjmean / VREF=0; BY group;
```

```
    TITLE 'Residuals';
```

```
RUN;
```

```
DATA voilib.resid_count; SET voilib.residuals;
```

```
    IF (-1.96<=resid<=1.96) THEN StudResidCount = 0;
```

```
    ELSE IF (resid ne.) THEN StudResidCount = 1;
```

```
    TITLE 'Residual Count (StudResidCount = 0 should be >=95%)';
```

```
RUN;
```

```
PROC FREQ; TABLE StudResidCount; RUN;
```

```
PROC LOESS DATA = voilib.residuals;
```

```
    ODS OUTPUT OutputStatistics=voilib.smresid;
```

```
    MODEL resid = epoch / smooth = 0.4;
```

```
    BY group;
```

```
RUN;
```

```

SYMBOL1 i=none c=green v=dot;
SYMBOL 2 i=none c=red v=dot;
SYMBOL 3 i=spline c=green v=none l=2 w=5;
SYMBOL 4 i=spline c=red v=none l=2 w=5;
LEGEND1 label=('Residuals');
LEGEND2 label=('Loess Curve');
axis1 order=(-10 to 10 by 2);

PROC GPLOT data=voilib.smresid;

    PLOT DepVar*epoch=group / vaxis = axis1 legend = LEGEND1;
    PLOT2 Pred*epoch=group / vaxis = axis1 legend = LEGEND2;
    TITLE 'Residuals by Time (should be no trend)';

RUN;

* plotting observed vs predicted *;
SYMBOL1 I=NONE C=GREEN V=TRIANGLE;
SYMBOL2 I=NONE C=RED V=HASH;
SYMBOL3 I=JOIN C=GREEN V=TRIANGLE L=2 W=4;
SYMBOL4 I=JOIN C=RED V=HASH L=2 W=4;
LEGEND1 LABEL=('Observed');
LEGEND2 LABEL=('Predicted');

PROC GPLOT DATA=voilib.residuals;

    PLOT adjmean*epoch=group / LEGEND = LEGEND1;
    /* VAXIS = -7 to 7 ; * To set axis limits if different for each
    plot; */
    PLOT2 pred*epoch=group / LEGEND = LEGEND2;
    * VAXIS = -7 to 7 ;
    FORMAT adjmean 8.1;
    TITLE 'Predicted vs Observed';

RUN;

```

The quantitative tests of normality in the proc univariate results give likelihoods of the each variable being normally distributed; for multivariable normality, all variables should be normally distributed, although this assumption is usually violated to some degree. The output of the UNIVARIATE procedure includes histograms of residuals (for assessing whether shape of distribution is approximately normal) and side-by-side box plots of residual distribution by group (Figure 3 and Figure 4).

Tests of influence

Further possible tests include influence diagnostics, to determine whether particular subjects have undue effect on the model. If undue bias due to individual subjects is suspected, the approach we suggest is to use the influence measures to investigate whether one or more subjects are consistently and substantially discrepant from the others. In SAS these tests may be run by adding the INFLUENCE

statement [e.g., after in PROC MIXED after “RESIDUAL” add “INFLUENCE (EFFECT=subject ITER=5)” and ODS SELECT INFLUENCE]; note that this requires SAS version 9.1 or greater. This test produces the restricted likelihood distance, PRESS statistic and Cook’s distance (known as Cook’s D) for each subject. If a discrepant subject is found, an attempt is made to determine whether there is some reason for this. If a reason can be found, then they might legitimately be excluded. A “sensitivity” analysis may be run (a comparison analysis including and excluding that patient) to see if there is a material difference in the conclusions or results of interest. The homogeneity of covariance matrices can also be assessed in a similar manner using the ESTIMATES option in the INFLUENCE statement. Although these influence diagnostics may be useful in some cases, they have less relevance in fMRI analyses due to the controlled nature of typical experiments and are presented here as optional.

Options if diagnostic tests fail

If the residuals are non-normal but symmetrical around 0 then the model is probably robust. However, if the residuals are both non-normal and non-symmetrical (i.e., skewed) then estimate biases may appear. For a large number of patients or observations, results are typically robust to moderate amount of skew, but for smaller numbers the model is likely to be robust for no more than a small amount of skew. If the data are seriously skewed, the typical options include making transformations (log, power transformations, etc.), identifying and omitting unusual patients (see above), using another more appropriate analytical technique, introduce covariates into the model, or not analyze the data. For fMRI studies, omitting unusual patients and adding covariates are the most likely approaches. As mentioned earlier, if the findings from the model are considered of value, then we suggest researchers report them along with the diagnostics test results so that the reader can assess for themselves the validity of the results.

Additional limitations

RMANOVA has additional limitations which may be less applicable to fMRI data, but are detailed below for completeness. The constraints of the assumptions result in weaknesses of the approach, including:

- Although not usually relevant to fMRI studies, cases with missing data will be entirely deleted from the analysis, causing both conceptual and analytical difficulties. Imputation methods can be employed to circumvent this issue.
- Tests of within-subjects effects assume sphericity, but this was not directly evaluated. The data can be tested for sphericity, but this evaluation is complex within the proc mixed framework, and is not addressed here. In theory, Mauchly’s test of sphericity can be used, for example, when any within-subjects factor has three or more trials. (If the within-subject factor fails to meet the assumption of sphericity, then either the multivariate approach can be used or the univariate results can be adjusted using a correction factor, e.g., the Huynh-Feldt Epsilon correction method [the Greenhouse-Geisser method has been shown to be too conservative]).

Figure 4. Plot of residuals for one group.

- The choice of the within-subject correlation matrix form is dependent upon sphericity assumptions, and we suggest a compound symmetric matrix (“type = cs” option in proc mixed) to allow for different variance patterns across groups. If different variance patterns across time are suspected, then heterogeneous compound symmetry could be used (“type = csh” option). However, this process is extremely computationally intensive, and in tests we performed, made no notable difference to the fit of the model.

Finally, the SAS library voilib created with the above SAS code contains numerous tables of data including predicted values, observations, model fit statistics, and residuals. These tables can be examined further either within SAS or by exporting them to a graph package, such as Excel.

Implementation and results

Test data and subjects

We used fMRI data collected from two obstructive sleep apnea and two healthy control subjects as part of a pilot project; data are available online⁹. The data were chosen to illustrate the RMANOVA methodology, and do not constitute a sample that is sufficient to test scientific hypotheses. The Institutional Review Board of UCLA approved this study (IRB# 10-001012), which was in compliance with the Declaration of Helsinki; participants provided informed, written consent after the nature of the procedures was explained. Following a baseline scanning period, subjects performed a series of four respiratory challenges (30 s maximal inspiratory apnea), which were expected to elicit abnormal physiological responses in the patient group, based on earlier demonstrations of impaired neural responses to physiologic challenges^{10–13}. A standard fMRI whole-brain protocol with repetition time of 2.5 s was

Figure 5. Box plot of residuals for both groups.

implemented on a 3T Siemens Trio MRI scanner, and high-resolution T1-weighted anatomical scans were also collected (voxel size = $0.9 \times 0.9 \times 1$ mm). The fMRI data were preprocessed using SPM routines, including realignment, spatial normalization, and smoothing. Smoothed images were intensity normalized to minimize global effects. VOI derived from the “AAL” toolbox in SPM were used to extract timetrends from the processed data¹⁴, and eight were selected to illustrate a variety of patterns highlighted by the approach. For each VOI, the time-series were analyzed using RMANOVA across the groups of two subjects, with challenges combined. Accompanying SAS and Excel files for each VOI are included online with this publication (Data availability).

Dataset 1. Processed data, as well as the SAS code for running each VOI analysis

<http://dx.doi.org/10.5256/f1000research.8252.d117479>

Organised as subfolders for each VOI, labeled by figure section. The VOI itself is also included as a nifti file, and in a pdf file overlaid onto a normalized background.

Preprocessing was performed using MATLAB 7 (The Mathworks, Inc., Natick, MA, USA), SPM (<http://www.fil.ion.ucl.ac.uk/spm/>) and a custom detrending routine⁶. VOI were drawn using MRICroN software¹⁵, and RMANOVA was implemented in SAS 9.4¹⁷.

Evaluation of fMRI responses, within and between groups

A set of results is presented in **Figure 6**, illustrating a variety of statistical responses for eight VOI within the same fMRI dataset. This section describes the analytic steps needed to arrive at these results.

Residual tests of the RMANOVA assumptions. The tests described in Appendix C were performed. The predicted responses were similar to the observed responses. Distributional tests of the Studentized residuals tests determined that the data were not precisely normally-distributed, as determined by the statistical tests with the PROC UNIVARIATE procedure. However, over 95% of the residuals were within ± 1.96 , and visual inspection of the plots of residuals and predicted versus observed values showed only minor skew, and therefore the data were considered to be adequately modeled by the RMANOVA.

Patterns highlighted by RMANOVA. All significant effects were BOLD signal responses. To infer neuronal responses, these patterns should be deconvolved with an HRF; this approach was not applied here.

The Tukey-Fisher criterion for multiple comparisons was applied, and the highlighted VOI did show significant overall effects (**Figure 6A–H**).

The three VOI in **Figure 6A–C** would all be highlighted as having significant group differences in response using conventional SPM analyses. However, RMANOVA is one way to illustrate the manner in which the groups differ. **Figure 6A** illustrates a typical fMRI pattern of difference, with one group showing increased signal throughout the challenge, and the other group essentially no change. **Figure 6B** illustrates an opposite response between the

Figure 6. Average time trends for a control and patient group, averaged over four breath-hold challenges (30 s maximal inspiratory apnea), from eight VOI. VOI are from the Automatic Anatomical Labelling toolbox extension to SPM. Time-points of between and within-group significant responses relative to baseline period are indicated by color-coded "*" symbols. The graphs are in percent signal change relative to baseline, and each trace is a group/challenge average ($\pm$ between-subject SE).

groups, with one increasing and the other decreasing during the challenge period. **Figure 6C** shows a decrease in one group with no change in the other.

Figure 6D illustrates a simple case of a similar change in both groups; thus the within-group timepoints of difference are highlighted during the challenge period, corresponding to an equivalent decrease in OSA and control subjects. **Figure 6E** shows both groups changing in the same direction, but with a greater magnitude of response in the OSA group, thus leading to significant between-group differences across multiple time-points. **Figure 6E** also illustrates a pattern of changes that lasts over 30 seconds into the recovery period.

Transient differences are shown in **Figure 6F–H**. **Figure 6F** shows a pattern of initial increased response in the OSA group and sustained decreased response in the control group. However, unlike previous examples, the trends in the OSA group change 10–15 seconds into the challenge from an increase to a decrease below baseline. **Figure 6G** shows an opposite pattern in control subjects (initial decrease, latter increase), and in OSA two large peaks are identified as differing significantly both between group and within group relative to baseline. Finally, **Figure 6H** shows a switch between greater OSA signal at the start of the challenge followed by a lower signal during the second half of the task period. These examples illustrate the capability of RMANOVA to detect the timing of within and between group differences in fMRI data.

Discussion and conclusions

Advantages of RMANOVA

The results illustrate two advantages of RMANOVA: firstly, no prior model of expected response of either signal intensity or timing is assumed, and for the type of challenge shown here, a variety of significant response patterns was detected by RMANOVA. Secondly, once a significant effect is found, RMANOVA provides an objective, statistically rigorous assessment of the time when responses or differences occurred, and the precise response pattern. A group difference highlighted by a traditional SPM analysis does not differentiate between a group increase vs. no change, a group increase vs. a decrease, no change vs. a decrease, or a larger increase vs. a smaller increase.

The procedure allows analysis of multiple subjects within multiple groups. The within-group and between-group are all performed with one analysis, allowing the contributions of subject and group factors to be accounted for in the final results.

Note that if all expected responses to the fMRI task are “on/off” or boxcar in nature, then the advantages of RMANOVA compared with cluster analysis are diminished.

Another advantage lies in processing timetrends from VOI drawn in different locations across different subjects. Spatial normalization is only accurate to within several millimeters¹⁶, and therefore small brain structures, such as those in the brainstem cannot be accurately depicted on VOI across multiple spatially normalized scans.

By allowing a group analysis of timetrends from varying sites, RMANOVA can be more accurate compared with conventional fMRI analyses in depicting group responses within a particular structure. For regions such as the dorsal medulla, VOI analysis using RMANOVA is particularly helpful for determining group patterns.

Disadvantages of RMANOVA

Two disadvantages of RMANOVA relate to defining the VOI. The procedure requires manual definition of the VOI, perhaps for each subject, or at least for each structure of interest. A second disadvantage is that no whole-brain search is performed, and only *a priori* defined areas are studied. For this reason, we believe that a combination of a traditional SPM analysis with RMANOVA allows for a complete and powerful approach to analysis of fMRI data where the time-course of responses is of interest.

The approach does not consider or allow for variations in the shape of the HRF. A standard HRF may be used to deconvolve each time series, resulting in an inferred neural response time series, or a different HRF may be used for different VOI (to account for the spatial variation in HRF shape).

Temporal autocorrelation between repeated measures is not accounted for, which likely leads to some loss of power. Nevertheless, we have found the technique highlights many patterns of interest, and thus this limitation does not negate the usefulness of the technique.

Procedural disadvantages with the proposed method include the requirement for detrending images prior to analysis, the extraction of the VOI time-trends from the images, and in the example presented here, the use of SAS. The former requires computation routines which are typically not included in fMRI analysis software, and the latter requires routines for exporting the data, availability of software, and some familiarity with this software.

There are a number of assumptions underlying the RMANOVA method, which are fully described above along with their associated limitations. As with any statistical model, RMANOVA should be used with problems that match those assumptions.

Conclusions

Analyzing timetrends from VOI's using RMANOVA can highlight temporal patterns of response to a challenge not readily apparent using conventional model-based approaches. Our research group has used this method extensively to assess fMRI responses to physiological challenges with complex responses over periods of tens of seconds to minutes. RMANOVA allows insight into the precise timing of changes from baseline, and response differences between groups. For complex paradigms, the technique can be a useful addition to conventional whole-brain model-based approaches.

Data and software availability

F1000Research: Dataset 1. Processed data, as well as the SAS code for running each VOI analysis, [10.5256/f1000research.8252.d117479](https://doi.org/10.5256/f1000research.8252.d117479)¹⁷

Harvard Dataverse: Macey, Paul. 2016. Pilot fMRI of breath-hold in OSA, <http://dx.doi.org/10.7910/DVN/EZUMI9>⁹

Author contributions

PMM and RMH conceived the study; PMM and PJS designed the methodology; PMM, KEM and RMH collected the test data and performed analysis and evaluation; all authors contributed to the manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

Supported by the National Institutes of Health National Institute of Nursing Research NR 013693.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Macey KE, Macey PM, Woo MA, *et al.*: **fMRI signal changes in response to forced expiratory loading in congenital central hypoventilation syndrome.** *J Appl Physiol* (1985). 2004; **97**(5): 1897–907.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ogren JA, Macey PM, Kumar R, *et al.*: **Central autonomic regulation in congenital central hypoventilation syndrome.** *Neuroscience*. 2010; **167**(4): 1249–56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Macey PM, Wu P, Kumar R, *et al.*: **Differential responses of the insular cortex gyri to autonomic challenges.** *Auton Neurosci*. 2012; **168**(1–2): 72–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Penny W, Friston K: **Mixtures of general linear models for functional neuroimaging.** *IEEE Trans Med Imaging*. 2003; **22**(4): 504–14.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Friston KJ, Holmes AP, Poline JB, *et al.*: **Analysis of fMRI time-series revisited.** *Neuroimage*. 1995; **2**(1): 45–53.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Macey PM, Macey KE, Kumar R, *et al.*: **A method for removal of global effects from fMRI time series.** *Neuroimage*. 2004; **22**(1): 360–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Littell RC, Milliken GA, Stroup WW, *et al.*: **SAS System for Mixed Models.** Cary, NC: SAS Institute Inc.; 1996.
[Reference Source](#)
- Macey PM, Valderama C, Kim AH, *et al.*: **Temporal trends of cardiac and respiratory responses to ventilatory challenges in congenital central hypoventilation syndrome.** *Pediatr Res*. 2004; **55**(6): 953–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Pilot fMRI of breath-hold in OSA.** *Harvard Dataverse*. 2016.
[Publisher Full Text](#)
- Macey KE, Macey PM, Woo MA, *et al.*: **Inspiratory loading elicits aberrant fMRI signal changes in obstructive sleep apnea.** *Respir Physiol Neurobiol*. 2006; **151**(1): 44–60.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Macey PM, Macey KE, Henderson LA, *et al.*: **Functional magnetic resonance imaging responses to expiratory loading in obstructive sleep apnea.** *Respir Physiol Neurobiol*. 2003; **138**(2–3): 275–90.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Henderson LA, Woo MA, Macey PM, *et al.*: **Neural responses during Valsalva maneuvers in obstructive sleep apnea syndrome.** *J Appl Physiol* (1985). 2003; **94**(3): 1063–74.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Harper RM, Macey PM, Henderson LA, *et al.*: **fMRI responses to cold pressor challenges in control and obstructive sleep apnea subjects.** *J Appl Physiol* (1985). 2003; **94**(4): 1583–95.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, *et al.*: **Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.** *Neuroimage*. 2002; **15**(1): 273–89.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rorden C, Karnath HO, Bonilha L: **Improving lesion-symptom mapping.** *J Cogn Neurosci*. 2007; **19**(7): 1081–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Krishnan S, Slavin MJ, Tran TT, *et al.*: **Accuracy of spatial normalization of the hippocampus: implications for fMRI research in memory disorders.** *Neuroimage*. 2006; **31**(2): 560–71.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Macey P, Schluter P, Macey K, *et al.*: **Dataset 1 in: Detecting variable responses within fMRI time-series of volumes-of-interest using repeated measures ANOVA.** *F1000Research*. 2016.
[Data Source](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 04 May 2016

<https://doi.org/10.5256/f1000research.8876.r13203>

© 2016 Brett M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Matthew Brett

Henry H. Wheeler, Jr. Brain Imaging Center, University of California, Berkeley, Berkeley, CA, USA

This paper shows how to apply the SAS "MIXED" procedure to fMRI data from ROIs, presents some data from this application, and discusses some of the options and diagnostics that this command provides.

The idea is to do what other packages would call an FIR analysis. You first decide on a time-bin width, typically one scan, but sometimes more. Call the time bin width B seconds. Then the scan data is labeled according to how many time bin widths it is after some event onset up until a cutoff of N time-bins.

Usually $N * B$ is equal to the length of the event plus 20 seconds or so to let the hemodynamic response drop back to baseline. Then the fitting procedure calculates adjusted means for each of the N bins, giving an evoked response time-course that can simultaneously adjust for other effects in the data, fit in the regression model.

In what follows, I confess I know SPM much better than the other packages, so I'll mainly compare to SPM.

SPM is rather general, but it does have this capability. The standard procedure would be to fit an "FIR" model to each subject, giving an estimated signal for each time bin in beta images, with one beta image for each bin. You then take these time-bin beta images into your second-level model, where you can adjust for variance covariance effects such as correlation between time-bin values within subjects, different within-subject variance, and so on.

As you say, using these methods in SPM can be moderately hard, but I'd argue the complexity is comparable to switching languages from MATLAB to SAS and using SAS to calculate things like adjusted means, residuals and so on.

How would you expect the results of using the SPM interface (I guess within marsbar) to compare to the results of using the SAS methods you have here.

I was expecting to see you using either MATLAB or R to do this; MATLAB because SPM already requires MATLAB, and it's a general programming language, R because it has a variety of procedures for fitting repeated measures models (see for example <http://www.r-statistics.com/2010/04/repeated-measures-anova-with-r-tutorials/>), is free, open-source, and, in my experience, is more widely used than SAS in research communities, including neuroimaging.

I'm afraid I haven't used SAS since the 1980s, so it's hard for me to comment on the details of the SAS commands.

I found myself skip-reading the actual SAS commands, I don't know how useful they will be for someone not experienced with SAS. Maybe in-line comments would help?

At various points you recommend options for diagnostics and outlier detection. It would help to have more detail as to why you recommend those options over others.

Description of procedure, first paragraph "Proportional scaling" is a technical term (at least in SPM-land) for dividing each scan voxel value by the calculated in-brain mean across voxels for the same scan. That is now very rarely used SPM, and was never widely used for FMRI. The standard SPM scaling involves multiplying each scan in a run by the same scaling factor, that brings the mean of in-brain mean voxel values to 100.

Assumptions, diagnostic checks, and limitations, first paragraph:

Could you expand on "Violations of independence produce a non-normal distribution of the residuals"?

Options if diagnostic tests fail, first paragraph "but for smaller numbers the model is likely to be robust for no more than a small amount of skew.". Could you rephrase to something like "For small numbers, even a small amount of skew will make it unlikely that the model is robust.". For some reason I found your original sentence a bit difficult to parse.

Disadvantages of RMANOVA, third paragraph:

You mention that "Temporal autocorrelation between repeated measures is not accounted for". Could you elaborate? I guess you mean that repeated measures variance / covariance adjustments in MIXED are for the estimates across time bins, not for the autocorrelation between scans. In that case, are you expecting something like SPM to do better, given that it does adjust for autocorrelation at the first level?

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 01 Jul 2016

Paul Macey, UCLA School of Nursing, LA, USA

[Initial comments repeated for both reviewers]

Thanks to the reviewers for their thoughtful comments. We made extensive revisions to the paper, which should make this method more accessible. The major changes including adding R code for implementing RMANOVA and diagnostic tests, simplifying some of the SAS commands and formatting requirements, and clarifying the nature of time series for which the method is applicable.

We did investigate implementing the method in MATLAB, but found that approach to be cumbersome and limited. While we did not test the method in SPSS, our understanding is that such implementation is similar to SAS, and relatively straight forward. In R, RMANOVA is non-trivial, and we settled on a mixed model approach that is equivalent to the one presented in SAS, but with more limited options should the user wish to try alternative covariance options. Furthermore, in R there are many options for diagnostic plots and calculations, and we selected one series of commands and packages that implement all SAS outputs other than studentized residuals.

In proposing a simplified data format, we decided to expand the scope of the method to include physiologic signals other than fMRI, such as heart rate, pulse oximetry, and other continuously acquired signals. We have applied our RMANOVA to these other signals in several publications. We therefore changed the title of the paper to reflect this broader application.

We updated the "dataverse" repository to include example R and SAS code, and a formatted data file.

Specific responses to Matthew Brett's comments

The reviewer's comments are helpful in clarifying how this method fits with other available approaches.

- *This paper shows how to apply the SAS "MIXED" procedure to FMRI data from ROIs, presents some data from this application, and discusses some of the options and diagnostics that this command provides.*

The idea is to do what other packages would call an FIR analysis. You first decide on a time-bin width, typically one scan, but sometimes more. Call the time bin width B seconds. Then the scan data is labeled according to how many time bin widths it is after some event onset up until a cutoff of N time-bins.

Usually $N * B$ is equal to the length of the event plus 20 seconds or so to let the hemodynamic response drop back to baseline. Then the fitting procedure calculates adjusted means for each for the N bins, giving an evoked response time-course that can simultaneously adjust for other effects in the data, fit in the regression model.

In what follows, I confess I know SPM much better than the other packages, so I'll mainly compare to SPM.

SPM is rather general, but it does have this capability. The standard procedure would be to fit an "FIR" model to each subject, giving an estimated signal for each time bin in beta images, with one beta image for each bin. You then take these time-bin beta images into your second-level model, where you can adjust for variance covariance effects such as correlation between time-bin values within subjects, different within-subject variance, and so on.

As you say, using these methods in SPM can be moderately hard, but I'd argue the complexity is comparable to switching languages from MATLAB to SAS and using SAS to calculate things like adjusted means, residuals and so on.

An SPM FIR analysis could replicate the SAS implementation of our timetrend RMANOVA approach, and we have used FIR in single-group fMRI studies of physiologic challenges. However, the results of a multi-group FIR quickly become complex since there is a voxel-level map of significant responses at each time point during the challenge and recovery. Performing between-group comparisons with FIR requires quite large and complex design matrices and contrasts. The marsbar toolbox allows extraction of the timetrend for those contrasts at pre-determined VOI, but to indicate timepoints of significant within or between group responses, the user would still need to identify these manually from the bin contrast maps (at least that is my understanding).

We agree that SAS is a complex environment unfamiliar to most neuroimaging researchers. At the time we originally developed this approach, MATLAB and R did not have the necessary capabilities, but that has mostly changed. In our revised paper, we present an implementation in R. We evaluated MATLAB but found no reasonable way to implement the present type of RMANOVA. In practice, the method is automated, so users don't need to work within R or SAS other than running the analyses.

One advantage of SAS over R is the option to easily choose a variety of covariance matrices. This could be important for experiments where the variability changes over time. For example, the standard "compound symmetric" (cs) covariance matrix we propose assumes equal variability at all timepoints including before during and after a task, whereas, for example, the "compound symmetric heterogenous" (csh) covariance matrix allows for different variances (and hence standard error estimates) at each timepoint. Theoretically csh should allow for more accurate models, although we found no major differences for our physiologic experiments and in our example use cs due to the lower computation needs. The R implementation does not have the option to choose covariance matrices, but we used a random intercept for each subject, which is equivalent to cs.

- *How would you expect the results of using the SPM interface (I guess within marsbar) to compare to the results of using the SAS methods you have here.*

The principal advantage of our technique is that it gives simple results that match the fMRI time-trend patterns. If a timetrend is higher or lower than the baseline at a timepoint, the method will indicate whether that increase was significant or not via a timepoint-specific p-value. If the timetrends for two groups diverge at a timepoint, the method will indicate whether that group difference is significant, again with a timepoint-specific p value.

- *I was expecting to see you using either MATLAB or R to do this; MATLAB because SPM already requires MATLAB, and it's a general programming language, R because it has a variety of procedures for fitting repeated measures models (see for example <http://www.r-statistics.com/2010/04/repeated-measures-anova-with-r-tutorials/>), is free, open-source, and, in my experience, is more widely used than SAS in research communities, including neuroimaging.*

I'm afraid I haven't used SAS since the 1980s, so it's hard for me to comment on the details of the SAS commands.

I found myself skip-reading the actual SAS commands, I don't know how useful they will be for someone not experienced with SAS. Maybe in-line comments would help?

We have provided an R implementation, and added comments.

- *At various points you recommend options for diagnostics and outlier detection. It would help to have more detail as to why you recommend those options over others.*

We have added additional comments explaining the tests.

- *Description of procedure, first paragraph "Proportional scaling" is a technical term (at least in SPM-land) for dividing each scan voxel value by the calculated in-brain mean across voxels for the same scan. That is now very rarely used SPM, and was never widely used for FMRI. The standard SPM scaling involves multiplying each scan in a run by the same scaling factor that brings the mean of in-brain mean voxel values to 100.*

We have removed this comment.

Assumptions, diagnostic checks, and limitations, first paragraph:

- *Could you expand on "Violations of independence produce a non-normal distribution of the residuals"?*

We added some further explanation of independence violations.

- *Options if diagnostic tests fail, first paragraph "but for smaller numbers the model is likely to be robust for no more than a small amount of skew.". Could you rephrase to something like "For small numbers, even a small amount of skew will make it unlikely that the model is robust.". For some reason I found your original sentence a bit difficult to parse.*

We agree the original wording was confusing, and we have reworded according to this suggestion.

Disadvantages of RMANOVA, third paragraph:

- *You mention that "Temporal autocorrelation between repeated measures is not accounted for". Could you elaborate? I guess you mean that repeated measures variance / covariance adjustments in MIXED are for the estimates across time bins, not for the autocorrelation between scans. In that case, are you expecting something like SPM to do better, given that it does adjust for autocorrelation at the first level?*

We give more detail in this paragraph. Specifically, we note that if there is a high correlation between adjacent time points (that is, there is only slow variation), this limitation will be more prominent. If the response are not expected to change much over time, SPM would be a better approach. We try to capture this aspect with "variable responses" in the paper title.

Competing Interests: No competing interests were disclosed.

Reviewer Report 26 April 2016

<https://doi.org/10.5256/f1000research.8876.r13288>

© 2016 Durnez J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Joke Durnez

Poldrack Lab, Department of Psychology, Stanford University, Stanford, CA, USA

This paper aims to present a different approach to analyzing fMRI time trends. The idea is to average the signal overall trials, over subjects (similar to ERP-analysis) and look at the differences between groups and conditions and especially its interaction with time since onset. Because the trials are clustered within subjects, the data points are modeled using a mixed model.

The result is a figure with the average time course per group/condition, with a statistical test comparing the BOLD signal at each time point.

The main difference between this model and the model in the standard (frequentist, parametric mixed model) procedure is the specific within-subjects design matrix. The predictor over time in a standard analysis would be:

[0 0 0 1 1 1 0 0 0 1 1 1] o HRF

The predictor over time with the presented approach would be a categorical variable:

[0 0 0 1 2 3 0 0 0 1 2 3]

The model tested is then $BOLD = GROUP + TIME + GROUP*TIME$. Important: TIME is modeled as a categorical variable, not continuous. Modeling time as a categorical folder allows to test for each time point separately as a post-hoc test. The results of interest are the post-hoc tests of the interaction effect between group and time.

SOME COMMENTS

1. The focus of the paper is on the post-hoc tests. These are definitely interesting, but the post-hoc tests should only be performed if the interaction group*time is significant. It would be interesting to also have some more detail on the main model (e.g. main effect of time, what does it mean? what is the interaction?)
2. In my opinion the chunks of code make it hard to read through the paper. It would be easier if the methods section only describes the different choices/tests/parameters and the code is grouped in an appendix. Especially code that is just for the purpose of making variables or figures should not be in the main manuscript.
3. In general, I think SAS is a bad choice. Even though it has definitely one of the best mixed model implementations, it's not a typical language at all in neuroimaging. If you want people to use this procedure, consider having a script in R, matlab or python.
4. I have tried to reproduce the code with the data attached. However, the datasets have different names and paths. It's hard to derive what is what...
5. I very much liked the diagnostics, something that is most often lacking in neuroimaging studies.
6. I think the part about missing data could be removed. This problem is non-existing in fMRI.
7. In the results section: "all significant effects were BOLD signal responses". Poor choice of words. The outcome variable is the BOLD signal response, not the effect.
8. "the highlighted VOI did show significant overall effects": significant overall effects of group? of time? the interaction? all of them?
9. The interpretation given to the time series is the interpretation of a pattern. However, these tests are not tests for patterns. For example, the interpretation of the interaction on fig 6E is not significant on all time points but is interpreted as such. This door is wide open for over interpreting non-significant results. This is a problem with these point wise comparisons.
10. There are ways to take away the temporal autocorrelation (whitening), which is currently used in fMRI software. You should mention this when talking about the problem of temporal correlation.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 01 Jul 2016

Paul Macey, UCLA School of Nursing, LA, USA

[Initial comments both reviewers]

Thanks to the reviewers for their thoughtful comments. We made extensive revisions to the paper, which should make this method more accessible. The major changes including adding R code for implementing RMANOVA and diagnostic tests, simplifying some of the SAS commands and formatting requirements, and clarifying the nature of time series for which the method is applicable.

We did investigate implementing the method in MATLAB, but found that approach to be cumbersome and limited. While we did not test the method in SPSS, our understanding is that such implementation is similar to SAS, and relatively straight forward. In R, RMANOVA is non-trivial, and we settled on a mixed model approach that is equivalent to the one presented in SAS, but with more limited options should the user wish to try alternative covariance options. Furthermore, in R there are many options for diagnostic plots and calculations, and we selected one series of commands and packages that implement all SAS outputs other than studentized residuals.

In proposing a simplified data format, we decided to expand the scope of the method to include physiologic signals other than fMRI, such as heart rate, pulse oximetry, and other continuously acquired signals. We have applied our RMANOVA to these other signals in several publications. We therefore changed the title of the paper to reflect this broader application.

We updated the "dataverse" repository to include example R and SAS code, and a formatted data file.

Specific responses to Joke Durnez's comments

This paper aims to present a different approach to analyzing fMRI time trends. The idea is to average the signal overall trials, over subjects (similar to ERP-analysis) and look at the differences between groups and conditions and especially its interaction with time since onset. Because the trials are clustered within subjects, the data points are modeled using a mixed model.

The result is a figure with the average time course per group/condition, with a statistical test comparing the BOLD signal at each time point.

The main difference between this model and the model in the standard (frequentist, parametric mixed model) procedure is the specific within-subjects design matrix. The predictor over time in a standard analysis would be:

[0 0 0 1 1 1 0 0 0 1 1 1] o HRF

The predictor over time with the presented approach would be a categorical variable:

[0 0 0 1 2 3 0 0 0 1 2 3]

The model tested is then $BOLD = GROUP + TIME + GROUP*TIME$. Important: $TIME$ is modeled as a categorical variable, not continuous. Modeling time as a categorical folder allows to test for each time point separately as a post-hoc test. The results of interest are the post-hoc tests of the interaction effect between group and time.

We appreciate this simplified description of the technique; we have included a synthesis at the start of the Introduction.

SOME COMMENTS

1. *The focus of the paper is on the post-hoc tests. These are definitely interesting, but the post-hoc tests should only be performed if the interaction group*time is significant. It would be interesting to also have some more detail on the main model (e.g. main effect of time, what does it mean? what is the interaction?)*

The sequence is to test 1) global fit of the model; if significant then test 2) variable and interaction level fit; and if significant test 3) post hoc at individual time points. This sequence was not clear in the original version. We agree that post hoc assessments at time points are only valid for significant model-level effects of time (within group) or group*time (between group), and we have emphasized this point in the revised paper.

2. *In my opinion the chunks of code make it hard to read through the paper. It would be easier if the methods section only describes the different choices/tests/parameters and the code is grouped in an appendix. Especially code that is just for the purpose of making variables or figures should not be in the main manuscript.*

We have simplified the code and proposed input data format, which eliminates some of the more cumbersome and less relevant code. We did originally have the code in appendices, but this format did not fit the requirements of this journal. In the revised paper, we added comments and restructured some sections to improve readability.

3. *In general, I think SAS is a bad choice. Even though it has definitely one of the best mixed model implementations, it's not a typical language at all in neuroimaging. If you want people to use this procedure, consider having a script in R, matlab or python.*

We now provide an implementation in R, since this software is free and well supported in the statistical community. We did evaluate RMANOVA in MATLAB, but found no reasonable approach to replicate the implementation we propose.

While SAS is an environment principally familiar to statisticians, the implementation of the procedure and diagnostic tests is simpler and more comprehensive than R. For that reason, we have left the SAS code in the paper.

4. *I have tried to reproduce the code with the data attached. However, the datasets have different names and paths. It's hard to derive what is what...*

We simplified the input data format to be a text file with three or four columns. The original approach of importing from Excel was cumbersome, but with the more logical format users should find it easier to run the SAS or R commands.

5. *I very much liked the diagnostics, something that is most often lacking in neuroimaging studies.*

We appreciate the comment. We have added explanations, as noted in response to Matthew Brett.

6. *I think the part about missing data could be removed. This problem is non-existing in fMRI.*

As noted in response to Matthew Brett, we have expanded the paper to explain how this RMANOVA approach applies to non-fMRI physiologic data, where missing values are possible. We have clarified the limitation.

7. *In the results section: "all significant effects were BOLD signal responses". Poor choice of words. The outcome variable is the BOLD signal response, not the effect.*

We revised this introductory paragraph.

8. *"the highlighted VOI did show significant overall effects": significant overall effects of group? of time? the interaction? all of them?*

We revised the wording in this section.

9. *The interpretation given to the time series is the interpretation of a pattern. However, these tests are not tests for patterns. For example, the interpretation of the interaction on fig 6E is not significant on all time points but is interpreted as such. This door is wide open for over interpreting non-significant results. This is a problem with these point wise comparisons.*

The reviewer is correct to note RMANOVA is not identifying patterns, but rather effects at time points independent of each other. We have changed this terminology.

We have emphasized that the time points will only be assessed post hoc if the corresponding independent variable is significant within the model. If the group×time interaction is not significant, the between-group time points will not be considered. If the time variable is not significant, the within-group time points will not be considered.

10. *There are ways to take away the temporal autocorrelation (whitening), which is currently used in fMRI software. You should mention this when talking about the problem of temporal correlation.*

We have expanded this paragraph and included a note about whitening.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research