Localizing Areas with Significant Inter-individual Variation: Testing Variance Components in a Multi-level GLM

Martin A. Lindquist, Julie Spicer, Lauren Leotti, Iris Asllani, and Tor D. Wager

Department of Statistics, Columbia University

Department of Psychology, Columbia University

INTRODUCTION

Most analysis of multi-subject fMRI data is concerned with determining whether a parameter (or contrast) of interest exhibits a significant population effect. However, valuable information may also be obtained by testing whether there are significant individual differences *between* subjects in response to a task, i.e. whether the variance components in the model significantly differ from 0. This type of test would provide a way of selecting regions that show significant inter-individual variability for subsequent analyses that attempt to explain those individual differences (e.g., brain-behavior correlations with reported pain, anxiety, performance scores, etc.). This strategy can circumvent problems associated with massive multiple comparison testing in undirected searches. Currently, the standard battery of tests performed on fMRI data doesn't include inference on variance components. In this work we discuss a framework for performing such tests in a multi-subject fMRI study.

METHODS

Assume a standard two-level GLM, with the first-level corresponding to subjects and the second to the group, i.e.

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \qquad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{V})$$

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}_G, \mathbf{U}_G)$$

for subject i=1,….m. In this analysis the variance component $\mathbf{U}_G$ is the parameter of interest, as it represents inter-individual differences in regression slopes. We assume it is a (q+1)×(q+1) matrix representing the covariance of the vector β_i. We use iterative generalized least-squares (IGLS) [1] to obtain maximum-likelihood estimates of the variance components, and restricted iterative generalized least-squares (RIGLS) [2] to obtain restricted maximum-likelihood estimates. Inference is performed on VCs using the likelihood ratio test (LRT), or restricted LRT (RLRT), which compares the full model (with q+1 random second-level components) to a reduced model excluding the component to be tested. The LRT can be expressed as:

$$LRT = -2\left[\log\left(L_{ML}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_0^*)\right) - \log\left(L_{ML}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^*)\right)\right]$$

where

$$\log\left(L_{ML}(\boldsymbol{\beta}, \boldsymbol{\beta}^*)\right) \propto -\frac{1}{2}\log\left|\boldsymbol{\Sigma}(\boldsymbol{\beta}^*)\right| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}(\boldsymbol{\beta}^*)^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

The parameters $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_0^*)$ and $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^*)$ are MLEs obtained by maximizing L_{ML} under the null and full model, respectively.

A valid RLRT for the variance components is obtained using ReML estimation. Here the general form of the test statistic remains the same, but L_{ML} is replaced by the ReML likelihood function

$$\log\left(L_{\mathrm{Re}ML}(\boldsymbol{\beta}, \boldsymbol{\beta}^*)\right) \propto -\frac{1}{2}\log\left|\mathbf{X}^T\boldsymbol{\Sigma}(\boldsymbol{\beta}^*)^{-1}\mathbf{X}\right| - \frac{1}{2}\log\left|\boldsymbol{\Sigma}(\boldsymbol{\beta}^*)\right| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}(\boldsymbol{\beta}^*)^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Null distributions

Standard statistical theory states that the asymptotic distribution of the LRT is χ^2_k where *k* is the difference in the number of parameters included in the full and reduced models. However, this result only holds under certain regularity assumptions; one being that the null hypothesis (H_0) does not lie on the boundary of parameter space, which is violated here (because the null is that inter-individual differences are zero). Hence, results from classical likelihood based inference are not directly applicable and modifications are required.

As an alternative Stram [3] suggested using a 50:50 mixture of χ^2 distributions with q+1 and q degrees of freedom to approximate the null distribution. Crainiceanu and Ruppert [5] derived an algorithm ("CR" in Simulations) to simulate the finite null distribution of the LRT, which gives more exact results than the mixture distribution. However, their results are only valid for models with a single between-subject variance component (q=0).

REFERENCES

[1] Goldstein, Biometrika 1986
[2] Goldstein, Biometrika 1989
[3] Stram, Biometrics 1994
[4] Morrell, Biometrics 1998
[5] Crainiceanu, JRSSB 2004
[6, 7] Wager et al., Neuroimage, in press

SIMULATIONS

A. Random Slope Model:

$$\mathbf{y}_i = \mathbf{1}\beta_0 + \mathbf{x}_i\beta_{1i} + \boldsymbol{\varepsilon}_i \quad \text{for i=1,….m}$$

Model

$$\beta_{1i} \sim N(\beta_1, \sigma_1^2)$$

$$\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2\mathbf{I}_n)$$

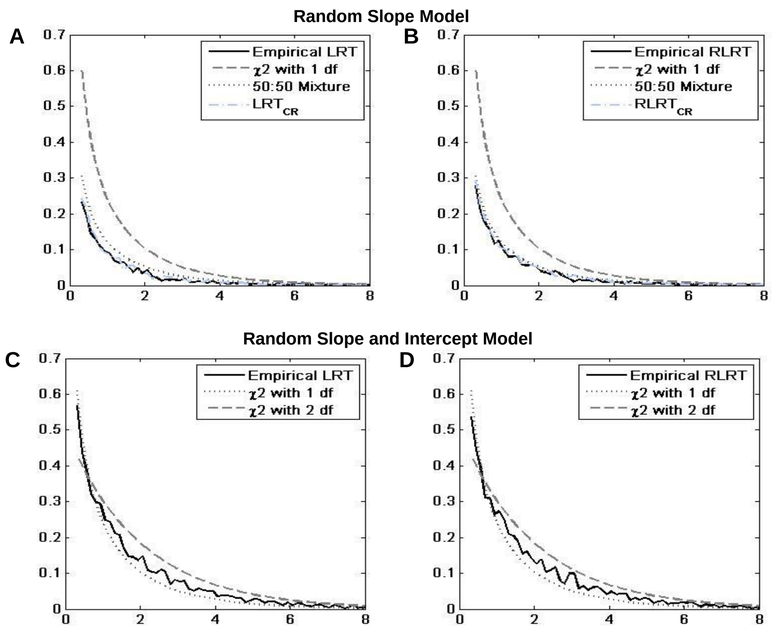
Test

$$H_0\colon\ \sigma_1^2 = 0$$

$$H_a\colon\ \sigma_1^2 > 0$$

Simulation I

Compare approximations of the null distribution of the LRT and RLRT statistic (50:50 mixture and CR) with the empirical null distribution for both the random slope and random intercept and slope models.



Panel A shows the empirical null distribution of the LRT statistic computed assuming the random slope model. This distribution acts as our ground truth in this simulation. The distribution used in standard likelihood ratio testing (χ^2_1) is overly conservative. The 50:50 mixture distribution also appears to be somewhat conservative. However, the approximation of the finite sampling distribution (LTR_{CR}) appears to accurately reflect the behavior in the tail of the distribution.

Panel B shows equivalent results for the RLRT. Here the difference between the 50:50 mixture distribution and $RLTR_{CR}$ is less pronounced.

Panels C and **D** show equivalent results for the random intercept and the slope model. The distribution used in standard likelihood ratio testing is overly conservative, while a 50:50 mixture accurately reflects the behavior in the tail of both distributions. Note that LTR_{CR} and $RLTR_{CR}$ are not defined when there are more than one between-subject VCs.

B. Random Intercept and Slope Model:

$$\mathbf{y}_i = \mathbf{1}\beta_{0i} + \mathbf{x}_i\beta_{1i} + \boldsymbol{\varepsilon}_i \quad \text{for i=1,….m}$$

$$\begin{pmatrix}\beta_{0i}\\ \beta_{1i}\end{pmatrix} \sim N\left(\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}, \begin{pmatrix}\sigma_0^2 & \sigma_{0,1}^2\\ \sigma_{0,1}^2 & \sigma_1^2\end{pmatrix}\right)$$

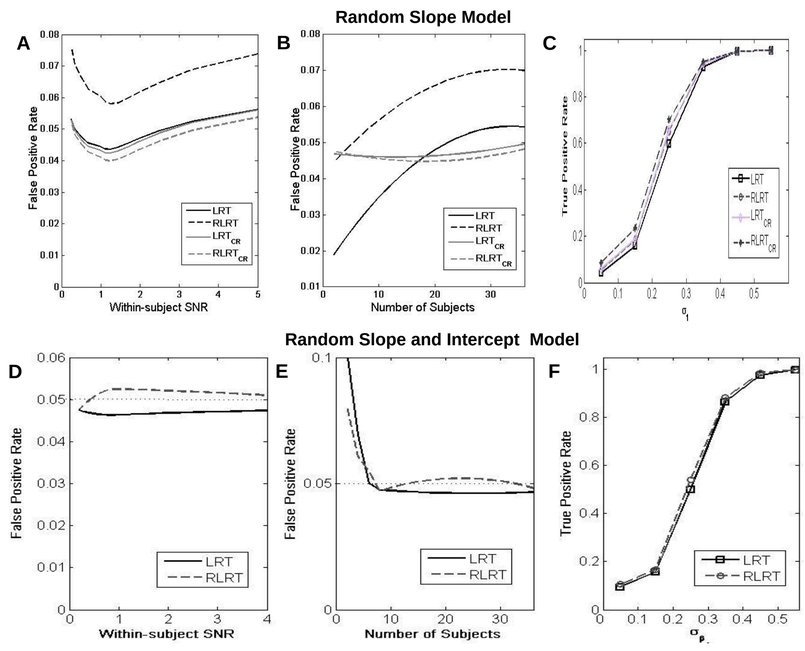
$$\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2\mathbf{I}_n)$$

$$H_0\colon\ \sigma_1^2 = 0$$

$$H_a\colon\ \sigma_1^2 > 0$$

Simulation II

Study the false positive rate and power of the LRT and RLRT as a function of within-subject SNR and number of subjects.



Panel A shows that the proportion of false positives, when the LRT and RLRT is thresholded at the 5% level, as a function of within-subject SNR. Though both are close to the nominal level of 0.05, the results based on the LRT are slightly more conservative than those based on the RLRT which is consistent with the literature [4]. **In general, the size of the within-subject SNR does not significantly impact the results.**

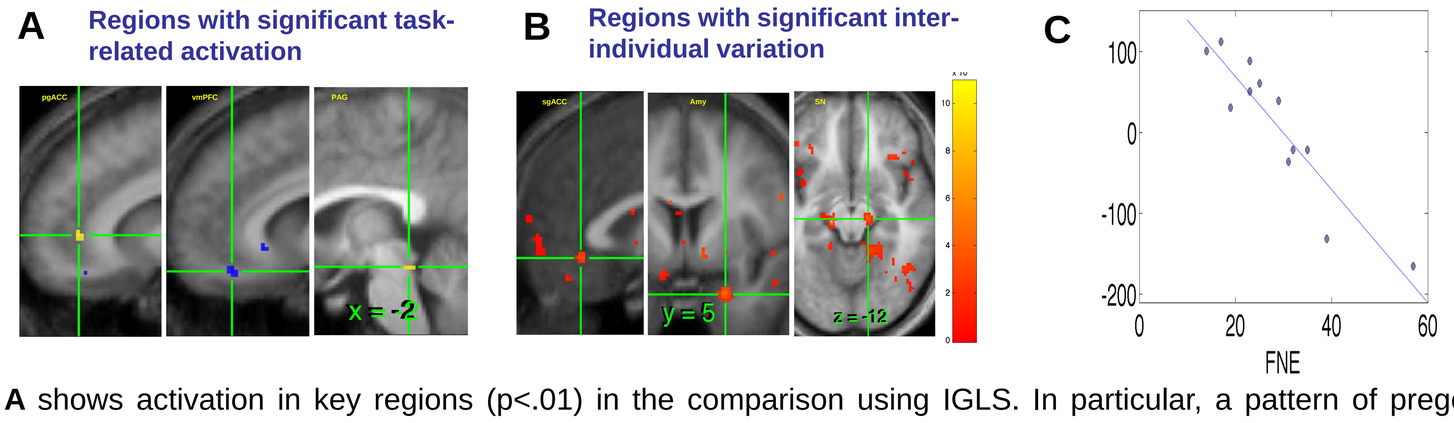
Panel B shows equivalent results for varying number of subjects. Both the LRT and RLRT give values close to the nominal values.

Panel C shows the true positive rate plotted as a function of σ_1. The RLRT gives a marginal increase in power compared to the LRT. **Hence, it appears ReML does not offer much benefit; a slight power increase, but also an increase in false positives.**

Panels D-F shows similar results for the random slope and intercept model. For small number of subjects (<10) there appear to be a large increase in the number of false positives. **Hence, one needs at least 6 subjects to fit a random slope and intercept model.**

RESULTS

The method was applied to cerebral blood flow measures collected with continuous arterial spin labeling (CASL) fMRI while performing a social evaluative threat task (n=12), allowing us to detect regions exhibiting significant individual differences. Participants rested for 5.3 min, and then silently prepared a speech to be given before an audience for 2.4 min. Data shown are for the contrast [Speech Preparation – Pre-stress Baseline].



Panel A shows activation in key regions (p<.01) in the comparison using IGLS. In particular, a pattern of pregenual cingulate increases (right), ventromedial prefrontal decreases (center), and periaqueductal gray (PAG) activity increases (right) replicates findings from two recent papers on the same task [6, 7].

Panel B shows areas with significant between-subjects variability in the same comparison. These regions are candidates for testing with explanatory variables.

Panel C shows that the sgACC region identified as showing high inter-individual variability (shown in B, left) was correlated with scores on trait Fear of Negative Evaluation (FNE; r=-.86, p<.002), which predicts physiological responses to social threat in other studies. **For more results see Poster 192 M-PM.**