Group Analysis with Both Individual Effect and Within-Subject Variability

Gang Chen, Ziad S. Saad, and Robert W. Cox
Scientific and Statistical Computing Core, National Institute of Mental Health
National Institutes of Health, Department of Health and Human Services, USA

1. Introduction

Data analysis in FMRI could be performed in one grand multilevel model that incorporates both within- and cross-subjects variabilities. However, it is due to the presence of both variance components that such a nonlinear model makes computation currently impractical, forcing analysts to adopt a two-stage approach: individual analysis with a time series regression model that explores activations in individual brains, and group analysis that integrates the results from all subjects.

Although it's been suggested that a frequentist approach in group analysis is unavailable in estimating variance components [1], we present here a computationally more economical method in a classical context than the Bayesian approach [1,2], while achieving equally valid inference by considering the effect precision from each subject. Our algorithms involve voxel-wise iterative schemes, and the computational cost is relatively low despite such an approach is considered computationally unmanageable [3].

2. Method

The conventional group analysis assumes the within-subject variability relatively small or roughly the same across subjects [4]. The corresponding model

$$\mathbf{b} = X^T \mathbf{a} + \mathbf{d}$$

implies that the effect estimate of interest (typically referred to as β value), $\mathbf{b}$, is accurate from individual level, with the precision information (estimated via the corresponding t-statistic) ignored at the group level, where the columns of design matrix X are either indicator

variables showing, for example, the group to which a subject belongs, or a subject-specific covariate such as IQ or behavioral data, **a** is group effect vector, and **d** is subject-specific error vector.

So much effort has been invested in modeling the temporal correlation in the FMRI regression model at individual level, leading to relatively more accurate statistical testing [5,6,7]. Such an effort should not stop there, and can further lead to more fruitful results by bringing the precision information about the effect of interest to group analysis. Here we consider a mixed-effects meta analysis (MEMA) model,

$$\mathbf{b} = X^T \mathbf{a} + \mathbf{d} + \mathbf{e}, \tag{1}$$

where **e** is the vector of sample errors from individual level, with a normal distribution with known variance. With model (1), the total variability in **b** is considered from two sources, within-subject and cross-subjects variability.

We estimate the cross-subject variance for **d** by maximizing the profile residual log-likelihood for the restricted maximum likelihood (REML) [8], estimate the group effect through weighted least squares (WLS).

To be able to deal with outliers in model (1), we assume a Laplace distribution for the subject-specific error term **d**, whose two heavier tails than normal distribution allow us to provide a more robust approach to alleviating the disturbing effect from outliers without arbitrary decision (*e.g.*, data removal) [9]. Although REML does not exist due to the introduction of Laplace distribution, our algorithm based on maximum likelihood estimate shows improved statistical power when outliers occur.

3. Results

A comparison of a paired-sample analysis shows that our model provides a more powerful detection of activation at most voxels compared to the conventional paired *t*-test (Fig. 1).

With heteroscedastic sampling precision incorporated in a MEMA model, we not only run a more accurate statistical testing, but also are able to estimate the heterogeneity measure (cross-subject variance) and test for its significance with Q-statistic. Furthermore, we can obtain the proportion of total variability that is within each subject, closely related to the popular concept of intraclass correlation (ICC). In addition, a Wald Z-test gives a significance test about the residuals of each subject, serving as another indicator of voxels or regions where a subject has exceptionally high or low effect/reliability. Combining the heterogeneity measure, the homogeneity Q-test, and the Wald test, one may be able to detect outlier regions or subjects, and to further investigate the possibility of including covariates or grouping subjects, thus fine-tuning the original model and increasing the statistical power.

Our program 3dMEMA is implemented in open source statistical language R [10] based algorithms discussed here, fully taking advantage of parallel computing with multi-core systems. In addition to gaining statistical power for the group effect of interest at most regions, this meta analysis approach leads to less spurious isolated voxels in the final result, alleviating some unnecessary FDR penalty during multiple testing correction.

As our algorithm for REML is reasonably efficient, the convergence is achieved within a few iterations at most voxels, leading to a runtime of a few minutes for a typical group analysis on a Mac OS X system with two 2.66 GHz dual-core Intel Xeon processors.

4. Conclusion

By considering a MEMA model with effect estimate and its precision from individual analysis, we adopt the conventional perspective, and use WLS method through maximizing the REML function. The program handles one-, two-, and paired-sample test types, and covariates are also allowed to control cross-subject variability. In addition to group effect estimate and its statistic, it also provides cross-subject heterogeneity estimate and a chi-square test for its significance, the percentage of within-subject variability relative to the total variance in the data, and a Z-statistic indicating the probability a subject is an outlier at a region.

Our new group analysis approach is computationally economical, and generally more powerful and valid than the conventional method of ignoring the effect estimate reliability from individual subjects. It is also relatively robust against outliers in the group data. The software is in open-source R and available for download.

References

[1] Woolrich, M. W. (2008), 'Robust group analysis using outlier inference', *NeuroImage*, 41(2): 286-301.

[2] Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson M., and Smith S. M. (2004), 'Multilevel linear modelling for FMRI group analysis using Bayesian inference', *NeuroImage*, 21(4):1732-47.

[3] Friston, K. J., Stephan, K. E., Lund, T. E., Morcom, A., and Kiebel, S. J. (2005). 'Mixed-effects and fMRI studies', *NeuroImage*, 24:244-252.

[4] Penny, W. D. and Holmes, A. J. (2007), 'Random effects analysis', In: *Statistical Parametric Mapping* (ed. Friston, K. *et al*.). Academic Press.

[5] Worsley, K. J., Liao, C., Aston, J., Petre, V., Duncan, G. H., Morales, F., Evans, A. C. (2002), 'A general statistical analysis for fMRI data', *NeuroImage*, 15:1-15.

[6] Woolrich, M., Ripley, B., Brady, J., and Smith, S. (2001), 'Temporal autocorrelation in univariate linear modelling of FMRI data', *NeuroImage*, 14(6):1370-1386.

[7] Kiebel S. J., and Holmes, A. P. (2007), 'The general linear model', In: *Statistical Parametric Mapping* (ed. Friston, K. *et al*.). Academic Press.

[8] Viechtbauer, W. (2005), 'Bias and efficiency of meta-analytic variance estimators in the random-effects model', *Journal of*
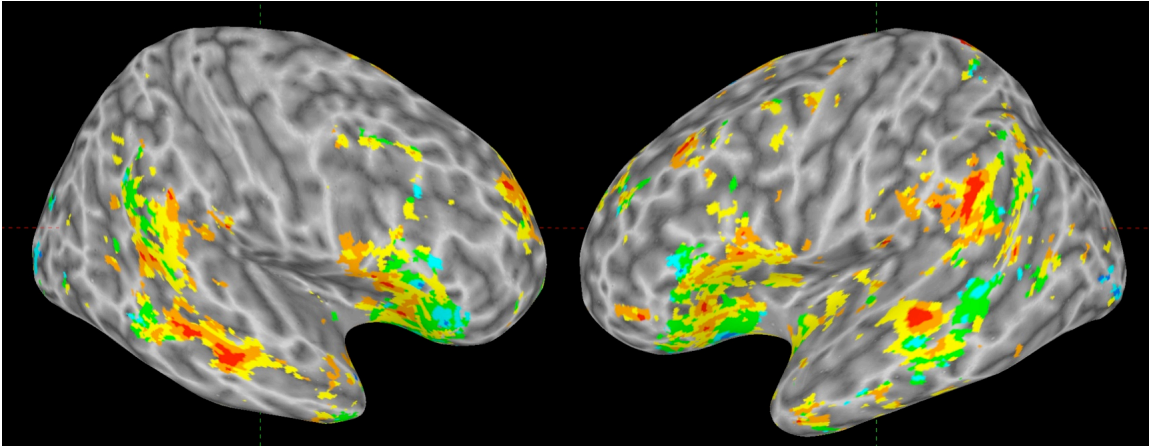
Educational and Behavioral Statistics, 30, 261-293.

[9] Demidenko, E. (2004), *Mixed Models: Theory and Applications*. Wiley-Interscience.

[10] R Development Core Team (2009), 'R: A language and environment for statistical computing', R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
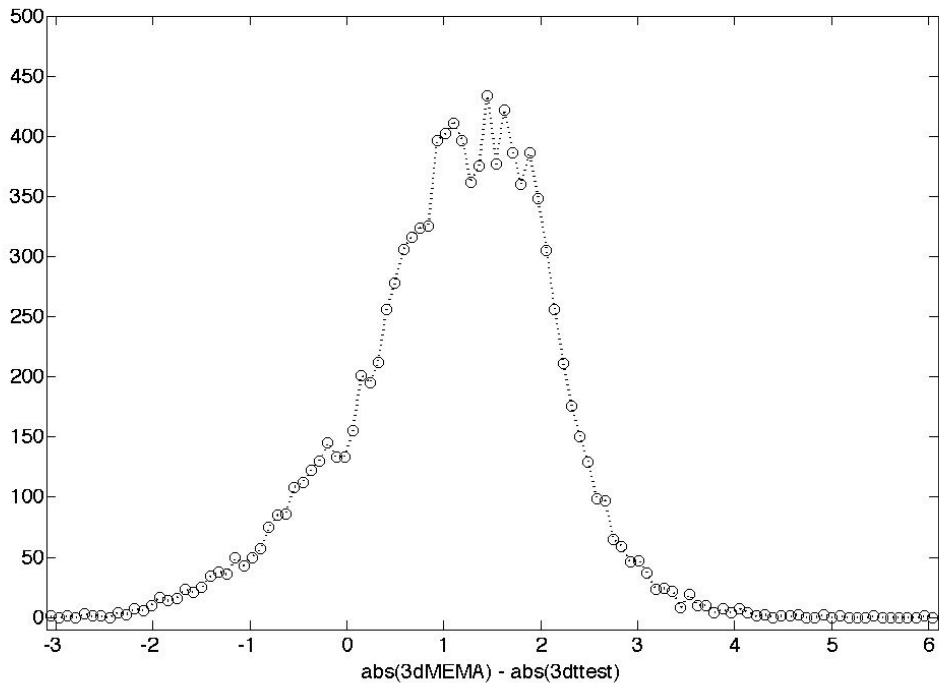
(A)

(B)

Figure 1. (A) Histogram of *t*-statistic difference: 3dMEMA gained power with a threshold of 2.0 for *t*(30) than the conventional approach with a paired t-test as shown by the majority of significant voxels at the right-hand side of the original at the x-axis. (B) Surface map of *t*-statistic difference: red >= 2.8, 1.7 ≤ orange < 2.8; 0.5 ≤ yellow < 1.7; -0.5 ≤ green < 0.5; blue ≤ -0.5) with a threshold of 2.0 for *t*(30). Courtesy of Vincent Costa, Univ. of Florida.