

# Random-Effects Analysis

W.D. Penny and A.J. Holmes

Wellcome Department of Imaging Neuroscience,  
University College London.

February 27, 2003

## 1 Introduction

In this chapter we are concerned with making statistical inferences from functional imaging studies involving many subjects. One can envisage two main reasons for studying multiple subjects. The first is that one may be interested in individual differences, as in many areas of psychology. The second, which is the one that concerns us here, is that one is interested in what is common to the subjects. In other words, we are interested in the stereotypical effect in the population from which the subjects are drawn.

As every experimentalist knows, a subject's response will vary from trial to trial. Further, this response will vary from subject to subject. These two sources of variability, within-subject (also called between-scan) and between-subject, must both be taken into account when making inferences about the population.

In statistical terminology, if we wish to take the variability of an effect into account we must consider the effect as a 'random effect'. In a 12-subject PET study, for example, we can view those 12 subjects as being randomly drawn from the population at large. The subject variable is then a random effect and, in this way, we are able to take the sampling variability into account and make inferences about the population from which the subjects were drawn. Conversely, if we view the subject variable as a 'fixed effect' then our inferences will relate only to those 12 subjects chosen.

The majority of early studies in neuroimaging combined data from multiple subjects using a 'Fixed-Effects' (FFX) approach. This methodology only takes into account the within-subject variability. It is used to report results as case studies. It is not possible to make formal inferences about population effects using FFX. Random-Effects (RFX) analysis, however, takes into account both sources of variation and makes it possible to make formal inferences about the population from which the subjects are drawn.

In this chapter we describe FFX and RFX analyses of a multiple-subject PET study. In section 2, we show how the analyses are implemented and in section 3 describe the underlying mathematical models. In neuroimaging, RFX is implemented using the computationally efficient 'summary-statistic' approach. We also show in section 3 that this is mathematically equivalent to the more computationally demanding maximum likelihood procedure.

## 2 Analysis of multi-subject data

Throughout this chapter we illustrate the different analysis methods using data from a PET study of verbal fluency. These data come from 5 subjects and were recorded under two alternating conditions. Subjects were asked to either repeat a heard letter or to respond with a word that began with that letter. These tasks are referred to as word shadowing and word generation and were performed in alternation over 12 scans and the order randomized over subjects. Both conditions were identically paced with one word being generated every two seconds. PET images were re-aligned, normalised and smoothed with a 16mm isotropic Gaussian kernel. <sup>1</sup>

### 2.1 Fixed-Effects Analysis

Analysis of multiple-subject data takes place within the machinery of the General Linear Model (GLM) as described in earlier chapters. However, instead of having data from a single-subject at each voxel we now have data from multiple subjects. This is entered into a GLM by concatenating data from all subjects into the single column vector  $Y$ . Commensurate with this augmented data vector is an augmented multi-subject design matrix <sup>2</sup>,  $X$ , which is shown in Figure 1. Columns 1 and 2 indicate scans taken during the word shadowing and word generation conditions respectively. Columns 3 to 10 indicate these conditions for the other subjects. The time variables in columns 11 to 15 are used to probe habituation effects. These variables are not of interest to us in this chapter but we include them to improve the fit of the model. The GLM can be written as

$$Y = X\beta + E \quad (1)$$

where  $\beta$  are regression coefficients and  $E$  is a vector of errors. The effects of interest can then be examined using an augmented contrast vector,  $c$ . For example, for the verbal fluency data the contrast vector

$$c = [-1, 1, -1, 1, -1, 1, -1, 1, -1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T \quad (2)$$

would be used to examine the differential effect of word generation versus word shadowing, averaged over the group of subjects. The corresponding t-statistic,

$$t = \frac{c^T \hat{\beta}}{\sqrt{\text{Var}[c^T \hat{\beta}]}} \quad (3)$$

where  $\text{Var}[\ ]$  denotes variance, highlights voxels with significantly non-zero differential activity. This shows the ‘average effect in the group’ and is a type of fixed-effects analysis. The resulting Statistical Parametric Map is shown in Figure 2(b).

It is also possible to look for differential effects in each subject separately using subject-specific contrasts. For example, to look at the activation from subject 2 one would use the contrast vector

$$c_2 = [0, 0, -1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T \quad (4)$$

<sup>1</sup>This data set and full details of the pre-processing are available from <http://www.fil.ion.ucl.ac.uk/spm/data>.

<sup>2</sup>This design was created using the ‘Multi-subject: condition by subject interaction and covariates’ option in SPM-99.

The corresponding subject-specific SPMs are shown in Figure 2(a).

We note that we have been able to look at subject-specific effects because the design matrix specified a ‘subject-separable model’. In these models the parameter estimates for each subject are unaffected by data from other subjects. This arises from the block-diagonal structure in the design matrix.

## 2.2 Random-Effects Analysis via Summary-Statistics

An RFX analysis can be implemented using the ‘Summary-Statistic (SS)’ approach as follows [3], [7].

1. Fit the model for each subject using different GLMs for each subject or by using a multiple-subject subject-separable GLM (as described in the last section). The latter approach may be procedurally more convenient whilst the former is less computationally demanding. For the purposes of RFX analysis the two approaches are mathematically identical as they will produce the same contrast images.
2. Define the effect of interest for each subject with a contrast vector. Each produces a contrast image containing the contrast of the parameter estimates at each voxel.
3. Feed the contrast images into a GLM that implements a one-sample t-test.

Modelling in step 1 is referred to as the ‘first-level’ of analysis whereas modelling in step 3 is referred to as the ‘second-level’. A balanced design is one in which all subjects have identical design matrices, and is a requirement for the SS approach to be valid.

If there are, say, two populations of interest and one is interested in making inferences about differences between populations then a two-sample t-test is used at the second level. It is not necessary that the numbers of subjects in each population be the same, but it is necessary to have the same design matrices for subjects in the same population ie. balanced designs at the first-level.

In Step 3, we have specified that only one contrast per subject be taken to the second level. This constraint may be relaxed if one takes into account the possibility that the contrasts may be correlated or be of unequal variance. This is discussed further in [5].

An SPM of the RFX analysis is shown in Figure 2(c). We note that, as compared to the SPM from the average effect in the group, there are far fewer voxels deemed significantly active. This is because RFX analysis takes into account the between-subject variability. If, for example, we were to ask the question ‘Would a new subject drawn from this population show any significant posterior activity?’, the answer would be uncertain. This is because three of the subjects in our sample show such activity but two subjects do not. Thus, based on such a small sample, we would say that our data do not show sufficient evidence against the null hypothesis that there is no population effect in posterior cortex. In contrast, the average effect in the group (in Figure 2(b)) is significant over posterior cortex. But this inference is with respect to the group of five subjects, not the population.

We end this section with a disclaimer, which is that the results presented in section 2, have been presented for tutorial purposes only. This is because

between-scan variance is so high in PET that results on single subjects are unreliable. For this reason, we have used uncorrected thresholds for the SPMs and, given that we have no prior anatomical hypothesis, this is not the correct thing to do [2] (see Chapter 14). But as our concern is merely to present a tutorial on the difference between RFX and FFX we have neglected these otherwise important points.

### 3 Variance components

This section is intended for the reader wishing to understand the statistical basis of the Summary Statistic approach to RFX and is written for the mathematically inclined. We also show how RFX and FFX differ.

In what follows  $\mathbf{E}[\cdot]$  denotes the expectation operator,  $\mathbf{Var}[\cdot]$  denotes the variance and we will make use of the following results. Under a linear transform  $y = ax + b$ , the variance of  $x$  changes according to

$$\mathbf{Var}[ax + b] = a^2 \mathbf{Var}[x] \quad (5)$$

Secondly, if  $\mathbf{Var}[x_i] = \mathbf{Var}[x]$  for all  $i$  then

$$\mathbf{Var} \left[ \frac{1}{N} \sum_{i=1}^N x_i \right] = \frac{1}{N} \mathbf{Var}[x] \quad (6)$$

For background reading on expectations, variance transformations and introductory mathematical statistics see [9].

#### 3.1 Random effects using maximum likelihood estimators

Underlying RFX analysis is a probability model defined as follows. We first envisage that the mean effect in the population (ie. averaged across subjects) is of size  $d_{pop}$  and that the variability of this effect between subjects is  $\sigma_b^2$ . The mean effect for the  $i$ th subject (ie. averaged across scans),  $d_i$ , is then assumed to be drawn from a Gaussian with mean  $d_{pop}$  and variance  $\sigma_b^2$ . This process reflects the fact that we are drawing subjects at random from a large population. We then take into account the within-subject (ie. across scan) variability by modelling the  $j$ th observed effect in subject  $i$  as being drawn from a Gaussian with mean  $d_i$  and variance  $\sigma_w^2$ . Note that  $\sigma_w^2$  is assumed to be the same for all subjects. This two-stage process is shown graphically in Figure 3.

Given a data set of effects from  $N$  subjects with  $n$  replications of that effect per subject, the population contrast is modelled by a two level process

$$\begin{aligned} d_{ij} &= d_i + e_{ij} \\ d_i &= d_{pop} + z_i \end{aligned} \quad (7)$$

where  $d_i$  is the true mean effect for subject  $i$  and  $d_{ij}$  is the  $j$ th observed effect for subject  $i$ . For the PET data the effect is a differential effect (the difference in activation between word generation and word shadowing). The first equation captures the within-subject variability and the second equation the between-subject variability.

The within-subject Gaussian error  $e_{ij}$  has zero mean and variance  $\text{Var}[e_{ij}] = \sigma_w^2$ . This assumes that the errors are independent over subjects and over replications within subject. The between-subject Gaussian error  $z_i$  has zero mean and variance  $\text{Var}[z_i] = \sigma_b^2$ . Collapsing the two levels into one gives

$$d_{ij} = d_{pop} + z_i + e_{ij} \quad (8)$$

The maximum-likelihood estimate of the population mean is

$$\hat{d}_{pop} = \frac{1}{Nn} \sum_{i=1}^N \sum_{j=1}^n d_{ij} \quad (9)$$

This estimate has a mean  $\text{E}[\hat{d}_{pop}] = d_{pop}$  and a variance given by

$$\begin{aligned} \text{Var}[\hat{d}_{pop}] &= \text{Var} \left[ \sum_{i=1}^N \frac{1}{N} \sum_{j=1}^n \frac{1}{n} (d_{pop} + z_i + e_{ij}) \right] \\ &= \text{Var} \left[ \sum_{i=1}^N \frac{1}{N} z_i \right] + \text{Var} \left[ \sum_{i=1}^N \frac{1}{N} \sum_{j=1}^n \frac{1}{n} e_{ij} \right] \\ &= \frac{\sigma_b^2}{N} + \frac{\sigma_w^2}{Nn} \end{aligned} \quad (10)$$

The variance of the population mean estimate contains contributions from both the within-subject and between-subject variance.

### 3.2 Fixed Effects

Implicit in FFX analysis is a single-level model

$$d_{ij} = d_i + e_{ij} \quad (11)$$

The parameter estimates for each subject are

$$\hat{d}_i = \frac{1}{n} \sum_{j=1}^n d_{ij} \quad (12)$$

which have a variance given by

$$\begin{aligned} \text{Var}[\hat{d}_i] &= \text{Var} \left[ \sum_{j=1}^n \frac{1}{n} d_{ij} \right] \\ &= \frac{\sigma_w^2}{n} \end{aligned} \quad (13)$$

The estimate of the group mean is then

$$\hat{d}_{pop} = \frac{1}{N} \sum_{i=1}^N \hat{d}_i \quad (14)$$

which has a variance

$$\begin{aligned}\text{Var}[\hat{d}_{pop}] &= \text{Var}\left[\sum_{i=1}^N \frac{1}{N} \hat{d}_i\right] \\ &= \frac{1}{N} \text{Var}[\hat{d}_i] \\ &= \frac{\sigma_w^2}{Nn}\end{aligned}\tag{15}$$

The variance of the fixed-effects group mean estimate contains contributions from within-subject terms only. It is not sensitive to between-subject variance. We are not therefore able to make formal inferences about population effects using FFX. We are restricted to informal inferences based on separate case studies or summary images showing the average group effect (eg. Figure 2(a) or Figure 2(b)).

### 3.3 Random effects using summary statistics

Implicit in the summary-statistic RFX approach is the two-level model

$$\begin{aligned}\bar{d}_i &= d_i + e_i \\ d_i &= d_{pop} + z_i\end{aligned}\tag{16}$$

where  $d_i$  is the true mean effect for subject  $i$ ,  $\bar{d}_i$  is the sample mean effect for subject  $i$  and  $d_{pop}$  is the true mean differential effect for the population.

The summary-statistic approach is of interest because it is computationally much simpler to implement than the full random effects model of equation 7. This is because it is based on the sample mean value,  $\bar{d}_i$ , rather than on all of the samples  $d_{ij}$ . This is important for neuroimaging as the images are so large.

In the first level we consider the variation of the sample mean for each subject around the true mean for each subject. The corresponding variance is  $\text{Var}[e_i] = \sigma_w^2/n$ , where  $\sigma_w^2$  is the within-subject variance. At the second level we consider the variation of the true subject means about the population mean where  $\text{Var}[z_i] = \sigma_b^2$ , the between-subject variance. We also have  $\text{E}[e_i] = \text{E}[z_i] = 0$ . Consequently

$$\bar{d}_i = d_{pop} + z_i + e_i\tag{17}$$

The population mean is then estimated as

$$\hat{d}_{pop} = \frac{1}{N} \sum_{i=1}^N \bar{d}_i\tag{18}$$

This estimate has a mean  $\text{E}[\hat{d}_{pop}] = d_{pop}$  and a variance given by

$$\begin{aligned}\text{Var}[\hat{d}_{pop}] &= \text{Var}\left[\sum_{i=1}^N \frac{1}{N} \bar{d}_i\right] \\ &= \text{Var}\left[\sum_{i=1}^N \frac{1}{N} z_i\right] + \text{Var}\left[\sum_{i=1}^N \frac{1}{N} e_i\right] \\ &= \frac{\sigma_b^2}{N} + \frac{\sigma_w^2}{Nn}\end{aligned}\tag{19}$$

Thus, the variance of the estimate of the population mean contains contributions from both the within-subject and between-subject variances. Importantly, both  $E[\hat{d}_{pop}]$  and  $\text{Var}[\hat{d}_{pop}]$  are identical to the maximum-likelihood estimates derived in section 3.1. This validates the summary-statistic approach. Informally, the validity of the summary-statistic approach lies in the fact that what is brought forward to the second-level is a *sample* mean. It contains an element of within-subject variability which when operated on at the second level produces just the right balance of within and between subject variance.

## 4 Discussion

We have shown how neuroimaging data from multiple subjects can be analysed using fixed-effects (FFX) or random-effects (RFX) analysis. FFX analysis is used for reporting case studies and RFX is used to make inferences about the population from which subjects are drawn. For a comparison of these and other methods for combining data from multiple subjects see [8].

In neuroimaging, RFX is implemented using the computationally efficient summary-statistic approach. We have shown that this is mathematically equivalent to the more computationally demanding maximum likelihood procedure. For unbalanced designs, however, the maximum-likelihood estimate of the population effect and its variance both change and the summary-statistic approach is no longer equivalent. The robustness of the summary-statistic approach to violations of these underlying assumptions is a topic covered in more detail in the following chapter.

For more advanced treatments of random effects analysis <sup>3</sup> see eg. [12]. These allow, for example, for subject-specific within-subject variances, unbalanced designs and for Bayesian inference [1]. For a recent application of these ideas to neuroimaging, readers are referred to Chapter 17 in which hierarchical models are applied to single and multiple subject fMRI studies. As groundwork for this more advanced material readers are encouraged to first read the tutorial in Chapter 13.

A general point to note, especially for fMRI, is that because the between-subject variance is larger than the within-subject variance your scanning time is best used to scan more subjects rather than to scan individual subjects for longer. In practice, this must be traded off against the time required to recruit and train subjects [11].

### 4.1 Further points

We have so far described how to make inferences about univariate effects in a single population. This is achieved in the summary statistic approach by taking forward a single contrast image per subject to the second level and the using a one sample t-test.

This methodology carries over naturally to more complex scenarios where we may have multiple populations or multivariate effects. For two populations, for example, we perform two-sample t-tests at the second level. An extreme example of this approach is the comparison of a single case study with a control

---

<sup>3</sup>Strictly, what in neuroimaging is known as random effects analysis is known in statistics as mixed effects analysis as the statistical models contain both fixed and random effects.

group. Whilst this may sound unfeasible, as one population has only a single member, a viable test can in fact be implemented by assuming that the two populations have the same variance.

For multivariate effects we take forward multiple contrast images per subject to the second level and perform an analysis of variance. This can be implemented in the usual way with a GLM but, importantly, we must take into account the fact that we have repeated measures for each subject and that each characteristic of interest may have a different variability. Methods for handling such cases are dealt with in Chapters 9 and 17.

As well as testing for whether univariate population effects are significantly different from hypothesized values (typically zero) it is also possible to test whether they are correlated with other variables of interest. In [10], for example, Ward et al. test to see whether task-related activation in the motor system correlates with age.

It is also possible to look for conjunctions at the second level. In [6], for example, Gottfried et al. test for areas that are conjointly active for pleasant, unpleasant and neutral odour valences. For a statistical test involving conjunctions of contrasts it is necessary that the contrast effects be uncorrelated. This can be ensured by taking into account the covariance structure at the second level.

The validity of all of the above approaches relies on the same criteria that underpin the univariate single population summary statistic approach. Namely, that the variance components and estimated parameter values are, on average, identical to those that would be obtained by the equivalent two-level maximum likelihood model.

## References

- [1] B.P. Carlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, 2000.
- [2] R.S.J. Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, and J.C. Mazziotta, editors. *Human Brain Function*. Academic Press USA, 1997.
- [3] L. Frison and S.J. Pocock. Repeated measures in clinical trials: An analysis using mean summary statistics and its implications for design. *Statistics in medicine*, 11:1685–1704, 1992.
- [4] K.J. Friston, D. Glaser, R. Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Applications. *Neuroimage*, 16:484–512, 2002.
- [5] D.E. Glaser, W.D. Penny, R.N. Henson, M.D. Rugg, and K.J. Friston. Correcting for non-sphericity in imaging data using classical and Bayesian approaches. *Neuroimage*, 13(6):S127, 2001.
- [6] J. A. Gottfried, R. Deichmann, J.S. Winston, and R.J. Dolan. Functional Heterogeneity in Human Olfactory Cortex: An Event-Related Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, 22(24):10819–10828, 2002.

- [7] A.P. Holmes and K.J. Friston. Generalisability, random effects and population inference. In *NeuroImage*, volume 7, page S754, 1998.
- [8] N.A. Lazar, B. Luna, J.A. Sweeney, and W.F. Eddy. Combining brains: a survey of methods for statistical pooling of information. *Neuroimage*, 16(2):538–550, 2002.
- [9] D.D. Wackerley, W. Mendenhall, and R.L. Scheaffer. *Mathematical statistics with applications*. Duxbury Press, 1996.
- [10] N.S. Ward and R.S.J. Frackowiak. Age related changes in the neural correlates of motor performance. *Brain, In Press*, 2003.
- [11] K. J. Worsley, C. H. Liao, J. Aston, V. Petre, G. H. Duncan, F. Morales, and A. C. Evans. A general statistical analysis for fMRI data. *NeuroImage*, 15(1), January 2002.
- [12] B.S. Yandell. *Practical data analysis for designed experiments*. Chapman and Hall, 1997.

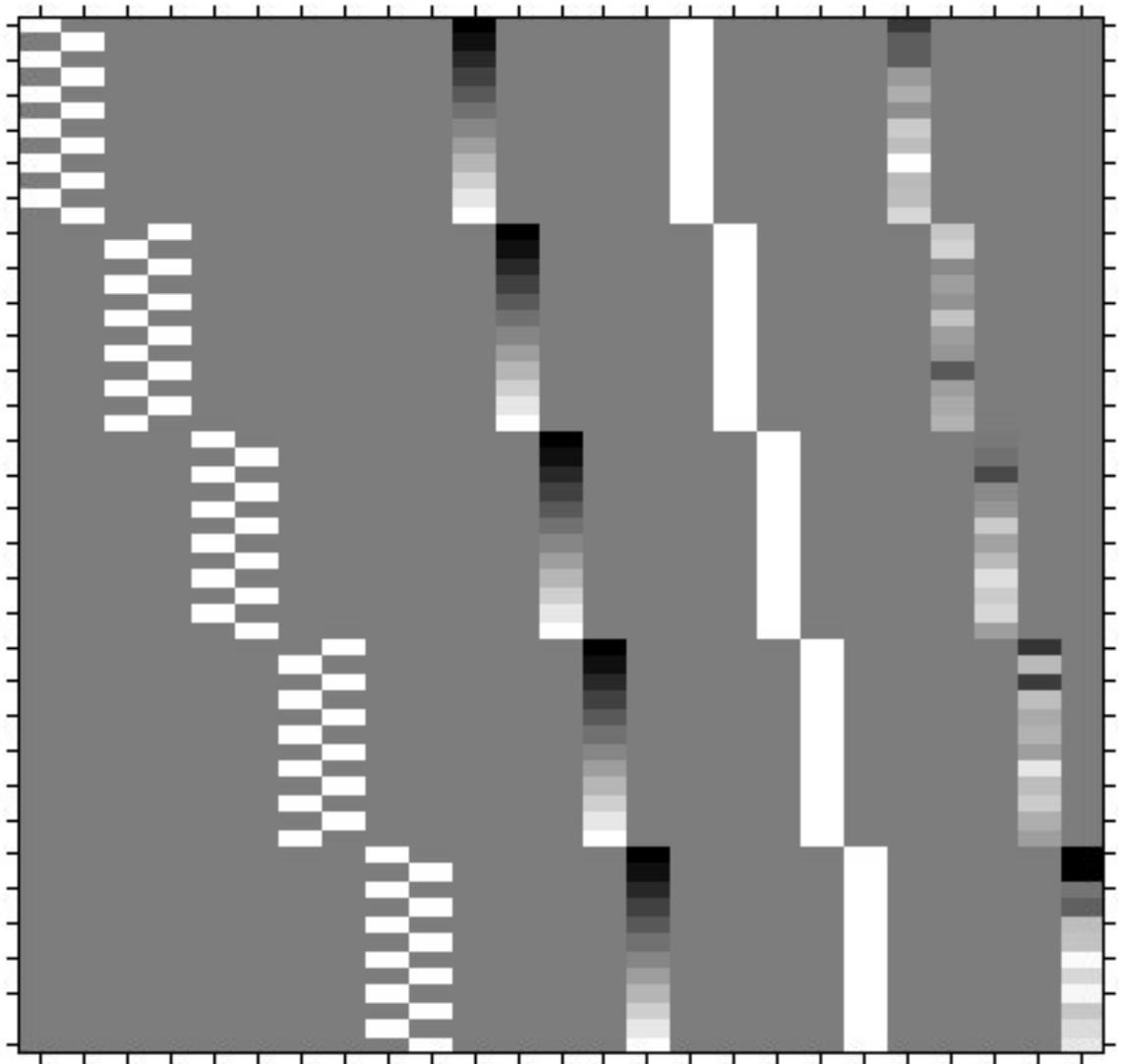


Figure 1: *Design matrix for five subject FFX analysis of PET data. There are 60 rows, 12 for each subject. The first ten columns contain indicator variables showing which condition (word shadowing or word generation) relates to which scan. Columns 11 to 15 contain time variables, columns 16 to 20 subject-specific offsets and the last 5 columns the global effect at each scan.*

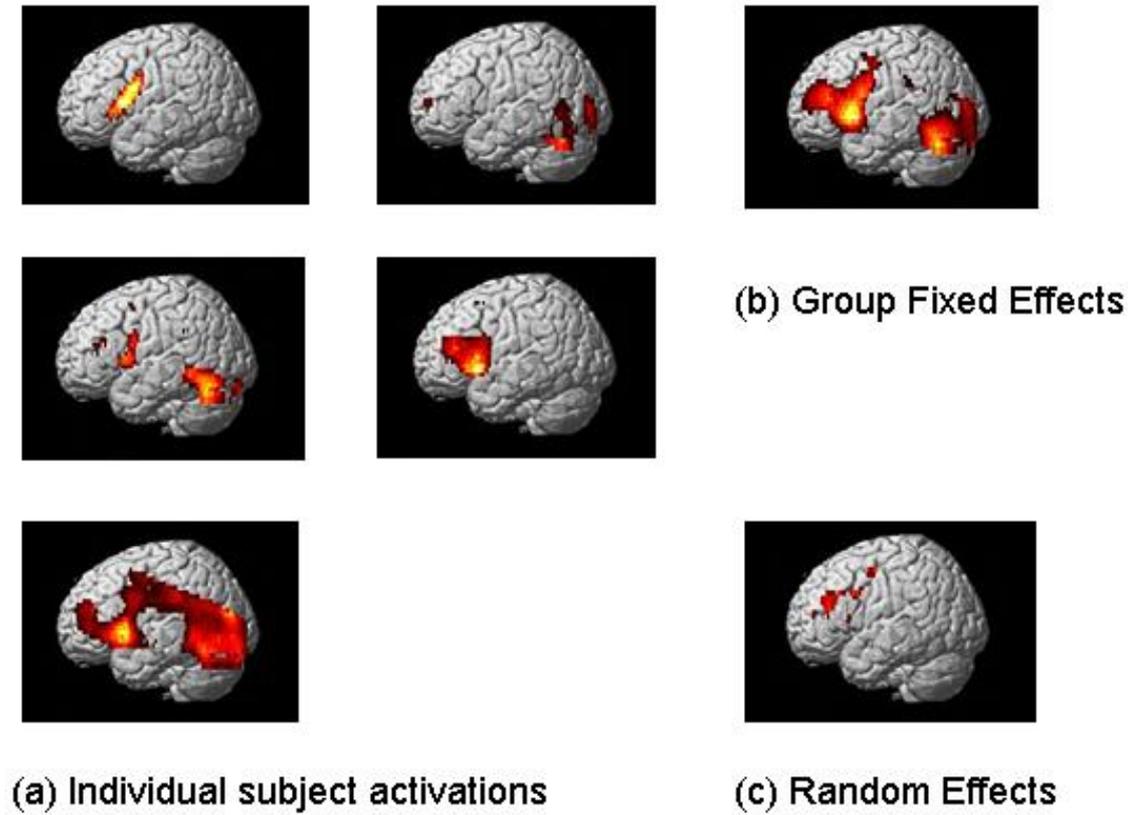


Figure 2: Analysis of PET data showing active voxels ( $p < 0.001$  uncorrected). The maps in (a) show the significance of subject-specific effects whereas map (b) shows the significance of the average effect over the group. Map (c) shows the significance of the population effect from an RFX analysis.

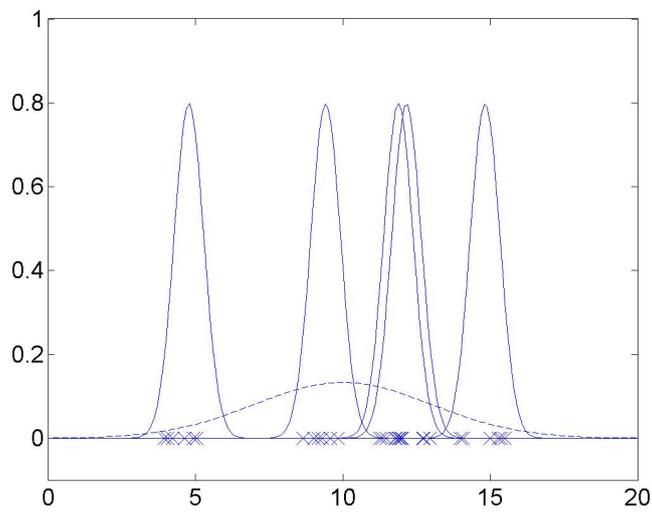


Figure 3: *Synthetic data illustrating the probability model underlying random effects analysis. The dotted line is the Gaussian distribution underlying the second level model with mean  $d_{pop}$ , the population effect, and variance  $\sigma_b^2$ , the between-subject variance. The mean subject effects,  $d_i$ , are drawn from this distribution. The solid lines are the Gaussians underlying the first level models with means  $d_i$  and variances  $\sigma_w^2$ . The crosses are the observed effects  $d_{ij}$  which are drawn from the solid Gaussians.*