HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2018 February 15.

Published in final edited form as:

Neuroimage. 2017 February 15; 147: 825–840. doi:10.1016/j.neuroimage.2016.08.029.

Untangling the Relatedness among Correlations, Part II: Inter-Subject Correlation Group Analysis through Linear Mixed-Effects Modeling

Gang Chen[*,a], **Paul A. Taylor**[a], **Yong-Wook Shin**[b], **Richard C. Reynolds**[a], and **Robert W. Cox**[a]

[a]Scientific and Statistical Computing Core, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, USA

[b]University of Ulsan College of Medicine, Department of Psychiatry, Asan Medical Center, South Korea

Abstract

It has been argued that naturalistic conditions in FMRI studies provide a useful paradigm for investigating perception and cognition through a synchronization measure, inter-subject correlation (ISC). However, one analytical stumbling block has been the fact that the ISC values associated with each single subject are not independent, and our previous paper (Chen et al., 2016) used simulations and analyses of real data to show that the methodologies adopted in the literature do not have the proper control for false positives. In the same paper, we proposed nonparametric subject-wise bootstrapping and permutation testing techniques for one and two groups, respectively, which account for the correlation structure, and these greatly outperformed the prior methods in controlling the false positive rate (FPR); that is, subject-wise bootstrapping (SWB) worked relatively well for both cases with one and two groups, and subject-wise permutation (SWP) testing was virtually ideal for group comparisons. Here we seek to explicate and adopt a parametric approach through linear mixed-effects (LME) modeling for studying the ISC values, building on the previous correlation framework, with the benefit that the LME platform offers wider adaptability, more powerful interpretations, and quality control checking capability than nonparametric methods. We describe both theoretical and practical issues involved in the modeling and the manner in which LME with crossed random effects (CRE) modeling is applied. A data-doubling step further allows us to conveniently track the subject index, and achieve easy implementations. We pit the LME approach against the best nonparametric methods, and find that the LME framework achieves proper control for false positives. The new LME methodologies are shown to be both efficient and robust, and they will be added as an additional option and settings in an existing open source program, 3dLME, in AFNI (http://afni.nimh.nih.gov).

[*]Corresponding author. gangchen@mail.nih.gov.

Introduction

Among various task-related FMRI experiments, one type of design ensures that a task is performed in a naturalistic way, in contrast to typical experiment designs in which a highly artificial task or condition is repeated with many trials to obtain a reliable effect estimate with adequate statistical power. Such naturalistic settings include movie watching, music or speech listening, and usually span the whole scanning session from the beginning to the end. The emphasis on natural scenes is placed so that more realistic and powerful activations can be detected (Hasson et al., 2004, 2008).

Unlike the statistical analysis with the typical task-related experiment, in which the activation in the brain is identified through one or more regressors that are associated with the explicit task timing, the investigator usually focuses on the synchronization or similarity between any pair of subjects. That is, one calculates the Pearson correlation between the EPI time series at the same location or voxel of the two subjects who underwent the same naturalistic-task scanning, which is termed as inter-subject correlation (ISC).

Two prototypical examples of ISC group analysis

We summarize briefly the background, framework, and structure of the ISC group analysis that was introduced in our previous nonparametric work (Chen et al., 2016), referred to hereafter as Part I, since the same concepts apply to the parametric modeling introduced here.

For one group with $n > 2$ subjects $S_1, S_2, \ldots, S_n$, the total $N=\frac{1}{2}n(n-1)$ unique ISC values $\{r_{ij}, i > j\}$ at each voxel form a symmetric ($r_{ij} = r_{ji}$) $n \times n$ positive semi-definite (PSD) matrix $\boldsymbol{R}^{(n)}$ with diagonals $r_{ii} = 1$ (Fig. 1). The Fisher transformed version $\boldsymbol{Z}^{(n)}$ (Fig. 1) is usually adopted during analysis so that parametric methods may be utilized under the Gaussianity assumption. The research of interest herein is focused on the ISC effect at the group level. Due to the symmetry of $\boldsymbol{R}^{(n)}$ and $\boldsymbol{Z}^{(n)}$, the group generalization can be made N elements in the lower (shaded gray in Fig. 1).

For two groups with n_1 and n_2 subjects ($n_1 + n_2 = n$), respectively, the matrices, $\boldsymbol{R}^{(n)}$ and $\boldsymbol{Z}^{(n)}$ would be similar to their counterparts with one group, with the additional consideration of a partitioned structure (Fig. 2). Again the focus of analysis is often on $\boldsymbol{Z}^{(n)}$, and then the effects of interest could reasonably be selected in several manners: 1) the ISC effect estimate for each group or within-group component (WGC), $\boldsymbol{Z}_{11}$ or $\boldsymbol{Z}_{22}$; 2) the direct group difference of ISC between the two groups, $\boldsymbol{Z}_{11}$ vs. $\boldsymbol{Z}_{22}$; 3) the between-groups component (BGC), $\boldsymbol{Z}_{21}$ (or $\boldsymbol{Z}_{12}$); or 4) the indirect contrasts $\boldsymbol{Z}_{11}$ vs. $\boldsymbol{Z}_{21}$, **and** $\boldsymbol{Z}_{22}$ vs. $\boldsymbol{Z}_{21}$.

ISC variance-covariance structure

Throughout this article, regular italic letters in lower (e.g., α) stand for scalars and random variables; boldfaced italic letters in lower ($\boldsymbol{a}$) and upper ($\boldsymbol{X}$) cases for column vectors and matrices, respectively. Major acronyms used in the paper are tabulated in the Appendix A. The notation z_{ij} is used in this paper with two meanings with respect to two subjects labeled as i and j: as the Fisher-transformed counterpart for an ISC value, and as a random variable

that can be conceptually thought of as representing the possible Fisher-transformed values[1]. A Fisher-transformed z-value is considered a sample or an instantiation of the associated random variable, and the specific usage should be evident in the respective context, even though no distinction is made in the notation.

Suppose that z_{ij} and z_{kl} are two z-values associated with the ISCs of r_{ij} and r_{kl} (i.e., $z_{ij} = arctanh(r_{ij})$) As in Part I, we denote the correlation between z_{ij} and z_{kl} that pivots around one subject (e.g., if $i = k$ or $i = l$) as ρ. That is, ρ is the "correlation of the correlation". To consider the group-wide set of ISCs, we further define $\mathbf{z} = vec(\{z_{ij}, i > j\})$ to be the vector of length N whose elements are the column-stacking of the lower triangular part of the matrix $\mathbf{Z}^{(n)}$ in Fig. 1 or its two-sample version. That is, $\mathbf{z}$ is the half-vectorization of $\mathbf{Z}^{(n)}$ excluding the main (or principal) diagonal: $\mathbf{z} = vech(\mathbf{Z}^{(n)}) \setminus diag(\mathbf{Z}^{(n)})$. The variance-covariance matrix of $\mathbf{z}$ can be expressed as the $N \times N$ matrix,

$$\sum^{(n)} = \sigma^2 \mathbf{P}^{(n)}, \quad (1)$$

where σ^2 is the variance of z_{ij}, $i > j$, and $\mathbf{P}^{(n)}$ is the correlation matrix that is composed of 1 (diagonals), ρ and 0 (Chen et al., 2016). As revealed in the variance-covariance structure (1) and discussed in Part I, the independence assumption typically required in the conventional parametric method, such as Student's t-test, does not hold for the group-wide dataset $\{z_{ij}, i > j\}$, with the nondiagnonal matrix $\mathbf{P}^{(n)}$ characterizing the degree to which the assumption of independence is violated (unless $\rho = 0$).

Previously, both parametric and nonparametric methods have been adopted in the literature and, for example, implemented in the Matlab package ISC Toolbox (Kauppi et al., 2014; https://www.nitrc.org/projects/isc-toolbox/). However, among the nonparametric approaches examined in Part I, the element-wise (EW) methods adopted in the literature mostly achieve poor or even unacceptable false positive rate (FPR) controllability, and should not be adopted in any ISC group analysis. In contrast, subject-wise bootstrapping (SWB) works relatively well for both cases with one and two groups (including the BGC $\mathbf{R}_{21}$) and group comparisons (including $\mathbf{R}_{11}$ / $\mathbf{R}_{22}$ versus $\mathbf{R}_{21}$), although its FPR controllability is sensitive to the correlation magnitude ρ (Part I). Lastly, subject-wise permutation (SWP) testing is virtually ideal for group comparisons, but performs poorly for the one-group scenario. However, even with the success of the nonparametric methods, they have the limitation of not allowing the analyst to include additional explanatory variables easily.

The ISC data structure is simply beyond the scope of the conventional parametric methods such as t-test, ANOVA and general linear model (GLM), which is evidenced by the adoptions of Student's t-test in the early days, with concomitant failures in controlling for FPR (Chen et al., 2016). Here, we start with the two prototypical examples, and partition each ISC value into the linear combination of three sources: unknown but fixed effects, the deviation/fluctuation for each of the two involved subjects from the fixed effects, and the

[1]Mathematically we may define the random variable z_{ij} here as a measurable function that maps two EPI time series to the real-valued space R^1.

residual or noise. This decomposition leads to relating the ISC data analysis to a special category of linear mixed- effects (LME) models, LME, with crossed random effects. We present both theoretical and practical features of utilizing these models, and then examine the FPR controllability and power for LME through simulations. A real experimental dataset is then used to assess its applications and performance. These performances are then compared to the best nonparametric approaches explored in Part I. We further explore the extendability of LME from the two prototypes, and discuss its interpretational power and other analytical benefits.

Methods

Effect partitioning of ISC data

In a typical FMRI analysis, the general interest at the group level is not about particular subjects nor their specific fluctuations, but rather focuses on hypotheses about the population effect or group difference. The situation and aim is the same with the ISC group analysis. The main challenge, after a recruitment of a large random sample, is to select an appropriate model to adequately characterize the variations embedded in the data while accounting for the intrinsic correlations of the basic quantitive unit, z_{ij} or r_{ij}. To construct a model with ISC data at the group level, we start with the simplest, one group case, by decomposing a given z_{ij} into multiple effects as follows:

$$z_{ij}=b_0+\theta_i+\theta_j+\varepsilon_{ij},\ i>j, \quad (2)$$

where b_0 is the fixed effect (an unknown constant), representing the group ISC effect; θ_i and θ_j are additive and independent random effects attributable to subjects i and j, respectively, that are deviations from (or adjustments to) b_0; and ε_{ij} is the residual or error term for each subject pair (i, j).

The inclusion of both fixed and random effects within the model lead to the general classification of (2) as being LME. We first describe the conceptual setup and analysis of this model in the standard LME context. While this approach is informative, we show that difficulties with both interpretation and practical implementation occur. We then proceed to examine the model within the specific category of LME modeling with crossed random effects (CRE), into which the relation of θ_i and θ_j allows it to be placed. We show how the CRE framework bypasses several issues which occur in the standard LME, and, in terms of its own practical considerations, how a balanced structure within the CRE model can be achieved. The latter framework is then tested on simulations and implemented on real FMRI, allowing comparisons to non-parametric methods from Paper I.

A brief description of ISC structure

The platform (2) offers an appealing feature that the correlation structure among $\{z_{ij}\}$ can be captured through the two random effects, θ_i and θj, and the error term ε_{ij}. With the basic assumption of Gaussian distributions for each individual term, $\theta_i, \theta_j \overset{iid}{\sim} G(0, \zeta^2)$ and $\varepsilon_{ij} \overset{iid}{\sim} G(0, \eta^2)$, we have a natural partitioning interpretation in the model (2): the resulting

variance of z_{ij}, σ^2, is additively composed of variance components from the random effects and residuals, $\sigma^2 = 2\zeta^2 + \eta^2$; and the factor of 2 in front of the cross-subject variability reflects the fact that the variability of each ISC comes from two independent sources, the two subjects. Furthermore, the covariance between two ISC values is conditionally describable, based on the pairs of subjects:

$$\begin{aligned} Cov(z_{ij}, z_{kl}) &= Cov(b_0+\theta_i+\theta_j+\varepsilon_{ij}, b_0+\theta_k+\theta_l+\varepsilon_{kl}) \\ &= \begin{cases} \sigma^2=2\zeta^2+\eta^2, & if\ the\ cardinality\ of\ \{i,j\}\cup\{k.l\}\ is\ 2\ (i.e., two\ distinct\ indices: i=k>j=l) \\ \zeta^2, & if\ the\ cardinality\ of\ \{i,j\}\cup\{k.l\}\ is\ 3\ (i.e., three\ distinct\ indices); \\ 0, & if\ the\ cardinality\ of\ \{i,j\}\cup\{k.l\}\ is\ 4\ (i.e., four\ distinct\ indices). \end{cases} \end{aligned}$$

Therefore, the correlation of a subject's z-values with two other subjects (e.g., z_{ij}, z_{jl}: cardinality of 3, $i > j > l$) can be expressed as the ratio of the cross-subjects variability relative to the total variance; specifically, with three subjects,

$$\rho = Corr(z_{ij}, z_{jl}) = \frac{Cov(z_{ij}, z_{jl})}{\sqrt{Var(z_{ij})\,Var(z_{jl})}} = \frac{\zeta^2}{2\zeta^2+\eta^2}. \quad (3)$$

This leads to the same overall correlation structure characterized earlier in $\boldsymbol{P}^{(n)}$ for $\{z_{ij}\}$ However, a benefit of the present modeling strategy, starting from the simple model (2), is this further description of ρ in terms cross- and within-subject variances, ζ^2 and η^2.

By inspections of (3), the allowed interval for the correlation ρ is seen to be

$$0 \le \rho = \frac{\zeta^2}{2\zeta^2+\eta^2} = \frac{\zeta^2}{\sigma^2} \le 0.5. \quad (4)$$

This is the same interval previously adopted for simulations in Part I using separate mathematical arguments based on the necessary and sufficient conditions for the positive semi-definiteness of $\boldsymbol{P}^{(n)}$; we note that the description in Part I also had contained an extra, vanishing interval $[-\frac{1}{2(n-2)}, 0)$ that was not considered of general interest, as the negative correlation between pairs of z-values from a given subject was not a realistic scenario. Furthermore the present derivation and description of ρ provides additional information on the boundary values in (4), $\rho = 0$ and 0.5, which correspond to the two extreme scenarios of having no cross-subjects variance ($\zeta^2 = 0$ and $\sigma^2 = \eta^2$) and no error term ($\eta^2 = 0$ and $\sigma^2 = 2\zeta^2$), respectively.

It is known that the Fisher-transformed Z-value for the correlation coefficient between two time series that are white noise approximately follows a Gaussian distribution $G(\frac{1}{2} ln \frac{1+r}{1-r}, \frac{1}{T-3})$ (Sheskin, 2004), where r is the Pearson correlation of two time series each having T time points. In brain regions where no BOLD signal is expected (e.g., white matter, particularly after regression of eroded WM signal and low order polynomials), we

would therefore expect the variance to be approximately $\sigma^2 = (T-3)^{-1}$; additionally, the ISC (and associated Z-) values are expected to be approximately 0, with group effect $b_0 = 0$ in the LME model (2) and cross-subjects variance $\zeta^2 = 0$, leading to an approximate within-subject variance

$$\sigma^2 = \eta^2 \approx \frac{1}{T-3}. \quad (5)$$

If a voxel's time series contains BOLD signal, then serial (temporal) correlation is expected to be present, and the within-subject variance η^2 would likely be higher than the value in (5), which effectively serves as an approximate lower bound for η^2 in the brain.

A preliminary ISC model at the group level

Through column-stacking of the lower triangular subset of the $\boldsymbol{Z}$ matrix in Fig. 1 into a vector $z_{N\times 1} = vec(z_{ij}, i > j)$, we can directly express the relationship (2) among the individual ISC data in a matrix format relevant for group analysis,

$$z = X\beta + K\pi + L\lambda + \varepsilon, \quad (6)$$

where $\boldsymbol{X}_{N\times N} = \boldsymbol{I}_N$, $\boldsymbol{\beta}_{N\times 1} = b_0\mathbf{1}_N$, $\boldsymbol{\pi}_{n\times 1} = \lambda_{n\times 1} = (\theta_1, \theta_2, \ldots \theta_n)^T$, $\boldsymbol{\varepsilon}_{N\times 1} = vec(\varepsilon_{ij}, i > j)$, and *vec* is the vectorization function. The model matrices $\boldsymbol{K}$ and $\boldsymbol{L}$ are structured to code which two indexed subjects i and j in $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$, respectively, added to the fixed effects for the outcome z_{ij}. While $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ are identical in this case, we express these contribution separately for modeling and implementation purposes, as discussed for them in the next paragraph and subsection.

In preceding discussions, the condition of having subject indices $i > j$ was simply imposed to focus on the lower triangular half of the symmetric matrix $\boldsymbol{Z}$. However, in the effect partition (2), this restriction has the practical consequence of leading to an unbalanced allotment of indices between the two random effect components. One way to address this imbalance would be to rearrange the index pairings through the specification of the model matrices for the two random effects; however, among the N elements in $\{z_{ij}, i > j\}$, each index pair (i, j) cannot always be evenly allotted between $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$: since there are $n-1$ occurrences of each index number, each subject index i cannot be evenly distributed between the two random effect components when the number of subjects, n, is even. Since it is not possible to generally balance the two model matrices for the random effects in this case, we simply structure the model matrices so that the first and second indices in each pair (i, j) are assigned to the first and second random-effects components $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$, respectively. This approach results in the following random-effect model matrices,

$$\begin{aligned} \boldsymbol{K}_{N\times n} &= (\boldsymbol{k}_1, \boldsymbol{k}_2, \cdots, \boldsymbol{k}_{n-1})^T, \\ \boldsymbol{L}_{N\times n} &= [\mathbf{1}_{n-1} \oplus \mathbf{1}_{n-2} \oplus \cdots \oplus \mathbf{1}_1, \mathbf{0}_{N\times 1}] \end{aligned} \quad (7)$$

where $\boldsymbol{k}_i = (\mathbf{0}_{(n-i)\times i}, \boldsymbol{I}_{n-i})^T$, and $\oplus$ is the direct sum operator for matrices.

Relationship of the preliminary ISC model with the conventional linear mixed-effects (LME) modeling

Linear mixed-effects (LME) modeling has been adopted in many disciplines, including neuroimaging (Chen et al., 2013; Bernal-Rusiel et al., 2013), because of its flexibility to deal with settings such as 1) when data are acquired from related (or "clustered") measuring units (e.g, the levels of a within-subject or repeated-measure factor), or 2) when missing data occur at random. For example, the conventional one-way, within-subject (or repeated-measures) ANOVA can be modeled under the LME platform as a special case having a random intercept. Directly relevant to the ISC context is the LME capability of specifying a correlation structure $\boldsymbol{P}^{(n)}$ for the ISC values that are associated with the same subjects, as each z_{ij} is generated by two subjects. Specifically, one can conceptualize that the ISC observations at the first level (i.e., the ISC pair) are clustered or even crisscrossed at the second level (i.e., subject) under the multilevel or hierarchical LME framework[2]. The second-level clusters (e.g., subjects) are often not of direct interest, but their random effects and correlation structures should be properly accounted for.

We could adopt the conventional, standard LME platform for the ISC data by merging the two random-effect components in the expression in (2) and (6). The vectorized expressions for the group would then be written as

$$z = X\beta + Y\theta + \varepsilon, \quad (8)$$

where the vector of random effects is $\boldsymbol{\theta}_{n\times 1} = \boldsymbol{\pi} = \boldsymbol{\lambda}$, and the model matrix that relates two subjects per outcome element in z is $\boldsymbol{Y}_{N\times n} = \boldsymbol{K} + \boldsymbol{L} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n-1})^T$, where $\boldsymbol{y}_i = (\mathbf{0}_{(n-1)\times(i-1)}, \mathbf{1}_{n-i}, \boldsymbol{I}_{n-i})^T$. In general, the components of $\boldsymbol{Y}$ are a column-wise subset of the fixed-effects model matrix $\boldsymbol{X}$, and the distributional assumptions are then that $\boldsymbol{\theta} \sim \boldsymbol{G}(\mathbf{0}, \boldsymbol{V})$, $\boldsymbol{\varepsilon} \sim \boldsymbol{G}(\mathbf{0}, \eta^2\boldsymbol{I})$, $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ are independent of each other (i.e., $\boldsymbol{\theta} \perp \boldsymbol{\varepsilon}$), and $\boldsymbol{V}$ is the variance-covariance matrix for the random effects $\boldsymbol{Y\theta}$.

However, while this standard LME, formulation is algebraically concise, the generalization is usually excessive, and the merger of the separate model matrices from (6) tends to be statistically "Procrustean," essentially masking the inherently independent relationship between the two random effect components. More importantly, the framework in (8) presents significant challenges and complications in calculating the multiple random effects $\boldsymbol{Y\theta}$. In contrast, it is the formulation at the subject level such as (2) that reveals the subject-specific details. Therefore, while broadly appealing at the outset, the consequences of the standard LME formulation lead to difficulty in both interpretation and implementation. In the next subsection we show how extensions of LME can address these difficulties.

[2]Note that we follow the statistics convention when adopting phrases such as "first level" and "second level" in the context of multilevel or hierarchical LME. Such phases are unrelated to the connotation of individual (or first level) and group (or second level) analysis commonly used in the neuroimaging community.

Theoretical generalization of the preliminary ISC model to the LME with CRE

One special characteristic of the LME model for ISC is being able to account for the combination of two random effect components: $\boldsymbol{\theta_i}$ and $\boldsymbol{\theta_j}$ in (2), or $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ in (6). That is, each observation z_{ij} is associated with the combination of two independent random effects, which are usually termed as *crossed*[3] random effects, in contrast to the conventional multilevel or hierarchical LME models in which the random effects are nested and correlated. LME models with crossed random effects have been explored and applied to psycholinguistic studies in recent years (e.g., Baaven et al., 2008), and for simplicity of notation here, we refer what is generally termed "linear mixed-effects models with crossed random effects" as "CREs," to clearly distinguish from standard LMEs. The generic formulation for two crossed random effects can be expressed at the individual element level (analogous to (2)) as

$$\boldsymbol{z}_{ij}=\boldsymbol{X}_{ij}\boldsymbol{\beta}+\boldsymbol{K}_i\boldsymbol{\pi}_i+\boldsymbol{L}_j\boldsymbol{\lambda}_j+\boldsymbol{\varepsilon}_{ij}, \quad (9)$$

where we note explicitly that, as in (2) the indicies *i* and *j* in (9) do not refer to vector/matrix elements, but instead to two measuring units (e.g., subjects, in the ISC case); the boldface quantities that they are labeling are vectors and matrices because, unlike in (2), this more general formulation may include multiple observations for given indices *i* and *j*, which track the levels of two crossed factors, such as subjects (or sites) and items (or time points).

Just as in the general LME formulation, the components of the model matrices $\boldsymbol{K_i}$ and $\boldsymbol{L_j}$for the two crossed random effects are either the same as or a column-wise subset of the model matrix $\boldsymbol{X_{ij}}$ for the fixed effects. Similarly, the distributional assumptions for (9) are that $\boldsymbol{\pi}_i \sim \boldsymbol{G}(0, \zeta_\pi^2 \boldsymbol{I}), \boldsymbol{\lambda}_j \sim \boldsymbol{G}(0, \zeta_\lambda^2 \boldsymbol{I})$, and $\boldsymbol{\varepsilon} \sim \boldsymbol{G}(\boldsymbol{0}, \eta^2 \boldsymbol{I})$, and also that $\boldsymbol{\pi}_i$, $\boldsymbol{\lambda}_j$ and $\boldsymbol{\varepsilon}_{ij}$ are independent of each other. The subscripts *i, j,* and *ij* in the model matrices $\boldsymbol{K_i}$, $\boldsymbol{L_j}$, and $\boldsymbol{X_{ij}}$ are presented in (9) for generic situations in which there may be different repetitions, covariate values, or missing data.

Our ISC formulation (2) can be subsumed and applied under the CRE platform (9) with only one observation per combination: $z_{ij} = z_{ij}$, $\boldsymbol{X}_{ij} = \boldsymbol{K}_i = \boldsymbol{L}_j = 1$, $\boldsymbol{\pi}_i = \theta_i$, $\boldsymbol{\lambda}_i = \theta_i$, and $\zeta_\pi^2 = \zeta_\lambda^2 = \zeta^2$. In other words the specific basic formulation (2) would still be stacked into a group formulation and expressed with the same model (6), but with the aim to analyze it as a CRE. The practical implementation of the model (9) as a CRE is described in the following subsection.

Considerations in numerically solving the LME models with CRE

Even though an LME system appears to be linear, in practice nonlinear solvers are required because the variances for the random effects are unknown a priori. For this reason, the LME

[3]The concept of crossed random effects is similar to that of factorial design in an ANOVA structure. As opposed to nested effects, for which each measuring unit (e.g., subject) of a factor (e.g., sex) occurs only under one level (e.g., female) of the factor, crossed effects are modeled when all the possible level combinations of the involved factors are present. For example, this is the case when there are deviations or calibrations θ_i and θ_j of subjects *i* and *j* from the fixed effects (except for the trivial case of $i = j$ or self correlation) in the LME framework (2)

model is typically solved through iteratively optimizing the corresponding restricted maximum likelihood (REML) function (e.g., Bates et al., 2015) with regard to both fixed- and random-effects parameters. This produces a best linear unbiased estimate (BLUE) for the fixed effects $\boldsymbol{\beta}$ in (8) and a best linear unbiased predictions (BLUP) for the random effects $\boldsymbol{\theta}$ in (8). Specifically for the CRE model (9), once the estimates for $\boldsymbol{\beta}$, ζ_{π}^2, ζ_{λ}^2, and $\boldsymbol{\eta}^2$ are obtained, omnibus *F*- and *t*-tests for fixed effects and their linear combinations can be constructed in the statistical conventional venue (Pinheiro and Bates, 2004). In practice, using any numerical solver introduces constraints on the user, such as input format, data structure, etc. Here, we implement the publicly available *R* package lme4 (Bates et al., 2015), and we discuss various practical issues and solutions within this framework; importantly, final validation of these methods are provided by the numerical simulations.

One practical subtlety is that the numerical solver for CREs is typically implemented (for example, in the *R* package lme4) with the general formulation (9), in which $\boldsymbol{\pi_i}$ and $\boldsymbol{\lambda_j}$ are not assumed to necessarily be the same and thus are solved for as two separate random effect components. We note that the use of the lme4 package will not allow combining the two random effects directly, as in (8). As a consequence, without the explicit imposition of $\zeta_{\pi}^2=\zeta_{\lambda}^2$, we will in general have two unequal variance estimates for ζ^2; specifically, the empirical estimate for the θ_i in $\boldsymbol{\pi}$ would likely be different for the same θ_i in $\boldsymbol{\lambda}$. Nevertheless, the fact with two unequal estimates of ζ^2 is not of a concern per se because the property of variance partitions means that we can practically utilize the average of the two estimates as the estimate for ζ^2.

Another aspect that needs consideration is the proper assignment for the degrees of freedom (DF) in an LME system. Statistical inferences through the conventional modeling approaches (Student's *t*-test, ANOVA, GLM) are typically made through some standard and exact distributions. The temptation of extending those well-behaved distributions to LME, although natural in real practice, cannot always be justified. Despite its great modeling flexibility and capability, one hindering (and hotly debated) aspect of LME is the daunting job of assigning DF or a *p*-value to each significance test. Unlike the conventional ANOVA or GLM, in which the effect partitioning is relatively straightforward and orthogonal most of the time, LME heavily relies on asymptotic properties, leading to a situation with no clearcut definition of DF, especially when a small or moderate sample size is combined with having 1) partially crossed random effects (here due to the index constraint $i > j$, or to the absence of combinations with $i < j$ from the upper triangular subset of $\boldsymbol{Z}$); 2) unbalanced structure; or 3) missing data. In addition, the shrinkage estimation through REML means that each individual random effect θ_i cannot be treated as an independent parameter. This DF difficulty is evidenced by the absence of degrees of freedom from the output in the function of lmer() in the *R* (R core team, 2016) package lme4 (Bates et al., 2015)[4]. One solution is to estimate the significance level based on the empirically sampled distribution (e.g., Markov chain Monte Carlo simulations) or nonparametric methods such as bootstrapping (Baaven et al., 2008). However, such methods are currently limited in applicability and computationally

[4]Also see http://glmm.wikidot.com/faq

too costly for the tens of thousands of voxels in neuroimaging data. Therefore here, we strive for an approach with reasonable computational time.

Practical LME formulation with fully crossed random effects and a balanced structure

To circumvent the impact of partially crossed random effect, here we loosen the index constraint of $i > j$ in (2) to $i \neq j$, thereby incorporating both the upper and lower triangular portions of the $\boldsymbol{Z}$ matrix in Fig. 1 into the formulation (8), which doubles the amount of ISC data (with redundancy) as input:

$$\begin{gathered}\boldsymbol{z}_{2N\times 1}=vec(z_{ij}, i\neq j),\\ \boldsymbol{X}_{2N\times 2N}=\boldsymbol{I}_{2N},\\ \boldsymbol{\beta}_{2N\times 1}=b_0\mathbf{1}_{2N},\\ \boldsymbol{\pi}_{n\times 1}=\boldsymbol{\lambda}_{n\times 1}=(\theta_1,\theta_2,\dots\theta_n)^T,\\ \boldsymbol{K}_{2N\times n}=(\boldsymbol{k}_1,\boldsymbol{k}_2,\cdots,\boldsymbol{k}_{n-1})^T,\\ \boldsymbol{L}_{2N\times n}=\boldsymbol{I}_n\otimes\mathbf{1}_{n-1},\\ \boldsymbol{\varepsilon}_{2N\times 1}=vec(\varepsilon_{ij}, i\neq j),\end{gathered} \tag{10}$$

where

$$\boldsymbol{k}_i=\begin{pmatrix}\boldsymbol{I}_{i-1} & \mathbf{0}_{(i-1)\times 1} & \mathbf{0}_{(i-1)\times(n-i)}\\ \mathbf{0}_{(n-i)\times(i-1)} & \mathbf{0}_{(n-i)\times 1} & \boldsymbol{I}_{n-i}\end{pmatrix}^T,$$

and $\otimes$ is the Kronecker product operator to matrices. A concrete example is that, with $n = 3$ subjects, we have

$$\boldsymbol{z}=\begin{pmatrix}z_{21}\\ z_{31}\\ z_{12}\\ z_{32}\\ z_{13}\\ z_{23}\end{pmatrix}, \boldsymbol{X}=\boldsymbol{I}_6, \boldsymbol{\beta}=b_0\begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix}, \boldsymbol{K}=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\\0&0&1\\1&0&0\\0&1&0\end{pmatrix}, \boldsymbol{L}=\begin{pmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{pmatrix}, \boldsymbol{\pi}=\boldsymbol{\lambda}=\begin{pmatrix}\theta_1\\ \theta_2\\ \theta_3\end{pmatrix}, \boldsymbol{\varepsilon}=\begin{pmatrix}\varepsilon_{21}\\ \varepsilon_{31}\\ \varepsilon_{12}\\ \varepsilon_{32}\\ \varepsilon_{13}\\ \varepsilon_{23}\end{pmatrix}$$

.

The data-doubling step essentially means that the whole data correlation matrix (except for the diagonals) is taken to the group analysis. One consequence of this is that the LME model (6)+(10) is symmetric, balanced[5] and fully crossed in the senses that: 1) for each index pair (i, j) or random effect pair (θ_i, θ_j) associated with z_{ij} there is a counterpart (j, i) or (θ_j, θ_i) because of the symmetric nature of the ISC matrix; 2) each subject is equally associated with the two random components, $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$, $n - 1$ times; and 3) even though $\boldsymbol{K}$ and $\boldsymbol{L}$ appear to be different in terms of matrix structure, there are exactly two 1s in each column for both $\boldsymbol{K}$ and $\boldsymbol{L}$, with each of the two matrices being a row-permuted version of the other. With this organization, the platform (6) can be easily implemented and the bookkeeping of indices is

[5]An LME system is balanced if the model matrices for the random effects are the same across subjects.

straightforward for the two crossed random effects. Furthermore, the balanced design with fully crossed random effect components between $\boldsymbol{K}$ and $\boldsymbol{L}$ should lead to exactly the same estimate for ζ^2, the shared variance of the two random effect components.

The CRE implementation for two groups of subjects closely follows that of the one group case. The basic platform (9) with one observation per combination can be formulated with fixed-effects model matrix $\boldsymbol{X}_{ij}$ of a row vector with two numbers, 1 (for the intercept) and another number coding the group membership, and again $\boldsymbol{K}_i = \boldsymbol{L}_j = 1$ (subset of $\boldsymbol{X}_{ij}$), $\boldsymbol{\pi}_i = \boldsymbol{\lambda}_i = \theta_i$, and $\zeta_\pi^2 = \zeta_\lambda^2$. Lastly, the modeling approach with (6) is directly applicable to any tests involving the BGC subset $\boldsymbol{Z}_{12}$ or $\boldsymbol{Z}_{21}$. Such tests include $\boldsymbol{Z}_{21}$, $\boldsymbol{Z}_1 - \boldsymbol{Z}_{21}$, $\boldsymbol{Z}_2 - \boldsymbol{Z}_{21}$, and $\boldsymbol{Z}_{21} - (\boldsymbol{Z}_1 + \boldsymbol{Z}_2)/2$. The data duplication in these cases translates to the utilization of both lower and upper counterparts (e.g., $\boldsymbol{Z}_{12}$ and $\boldsymbol{Z}_{21}$) as input.

Finally, with a system of inherently and fully crossed random effects having specifically balanced structures, we must address the issue of assigning the general concern about the degrees of freedom in a consistent manner. Firstly, to compensate for the potential inflation due to the data redundancy of including both lower and upper triangular subsets of $\boldsymbol{Z}$, which double the number of elements from N to $2N$, we make an adjustment from $2N - k$ to $N - k$ in the DF of the standard error, where k is the column rank of the fixed-effect model matrix. Secondly, as there are only n independent measuring units (subjects), we discount the nominal number of DF directly from the model from $2N - k$ for each t-test. The validity of our DF assignment scheme will be explored quantitatively through simulations in the section Simulations and Real Experiment Results.

Results: Simulations and Real Experiment

Simulations of group analysis with different testing methods

We performed simulations in a 2 × 4 × 6 × 3 factorial design with our focus on:

- **a.** 2 types of ISC group analysis: one- and two-sample (one and two groups, respectively);
- **b.** 4 sample sizes: 10, 20, 40 and 80 subjects in each group;
- **c.** 6 parameter values: Six ρ values were selected from the interval of [0, 0.5] with a step size of 0.1; and
- **d.** 3 testing methods: (1) subject-wise permutations (SWP) from Part I, (2) subject-wise bootstrapping (SWB) from Part I, and (3) CRE. For the two-sample scenario, four tests were performed: direct contrast between the two WGCs ($\boldsymbol{Z}_{11}$ vs $\boldsymbol{Z}_{22}$), the BGC effect of $\boldsymbol{Z}_{21}$, and indirect contrast of each WGC versus the BGC ($\boldsymbol{Z}_{11}$ vs $\boldsymbol{Z}_{21}$ and $\boldsymbol{Z}_{22}$ vs $\boldsymbol{Z}_{21}$).

To examine the FPR controllability and power attainment for each of the 2 × 4 × 6 × 3 scenarios, 5000 simulated single voxel datasets were generated, each containing

$N = \frac{1}{2}n(n-1)$ values drawn from an N-variate Gaussian distribution $G(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with the variance-covariance matrix $\boldsymbol{\Sigma}$ defined per the structure (1), where $\sigma^2 = 1$. For FPR

evaluation, $\boldsymbol{\mu_N} = \mathbf{0}_N$ for all cases. The same datasets were adopted for power assessment except that the mean was shifted to $\boldsymbol{\mu}_N = 0.5 \cdot \mathbf{1}_N^T$ (one group) or $\boldsymbol{\mu}_N = (\mathbf{0}_{N-N_2}^T, 0.5 \cdot \mathbf{1}_{N_2}^T)^T$ (two groups, where $N_2 = \frac{1}{2} n_2 (n_2 - 1)$ is the number of elements in $\boldsymbol{Z}^{(n)}$ for the second group). FPR and power for each scenario were estimated by counting the number of realizations out of the simulated 5000 datasets that reached the nominal significance level of 0.05.

Initial simulations with standard LME indicated that, without duplication of $z_{ij}(i > j)$, some of the tests (e.g., one group, $\mathbf{Z}_{11} - \mathbf{Z}_{22}$ for two groups, not shown here) were largely inflated in FPR when $\rho \geq 0.1$ (**FPR** ∈ [**0.1, 0.16**] with a nominal significance level of 0.05) and over-conservative when $\rho \approx 0$. Therefore, standard LME results are not considered further, and hereafter, we discuss only the simulation results with the duplication approach for LME/CRE.

The FPR and power estimates for the three methods are shown in Fig. 3, and they tend to follow a similar pattern: a conservative FPR control is usually associated with a lower performance in power. For the one-group case (upper row in Fig. 3), LME (green) in general shows better FRP controllability than SWB (blue). Specifically, the FPR for SWB (blue) monotonically increases as the correlation parameter ρ goes from 0 to 0.5 across all the sample sizes, while LME's FPR starts conservatively and then matches reasonably well with the nominal significance level once ρ reaches 0.1. When $\rho = 0$, SWB is too conservative: as noted in Part I, in this special case, the "bundling" of ISC values per subject, when they are actually independent of each other, leads to overly conservative identifications. When ρ is around 0.15, SWB is well-behaved in FPR; but when $\rho > 0.2$, SWB becomes moderately liberal. LME is conservative as well ρ is close to 0, but in contrast the severity quickly tapers off when the number of subjects increases, and the FPR performance is basically satisfactory when $\rho > 0.05$. In contrast, the power attained by LME and SWB is consistent with their respective FPR controllability. Specifically, SWP slightly outperforms LME in general (except when $\rho < 0.1$ with 10 subjects), but the differences vanish when the number of subjects gets close to 80. On the other hand, with two groups LME is slightly more powerful for all ρ values and all sample sizes than SWP.

For the case with two groups (second row in Fig. 3), SWP (black) is uniformly well-behaved in terms of FPR controllability across all sample sizes and across the whole range of ρ values. While LME is again slightly conservative when ρ is very close to 0, its FPR is within the 95% confidence band when $\rho \geq 0.1$ and there are 20 or more subjects in each group.

An important feature which is not available from the nonparametric SWB/SWP modeling is that, with LME, both the cross- and within-subject variances, ζ^2 and η^2, can also be estimated (Fig. 4) from the simulations. The effect partitioning guarantees that the sum of the variances equals the assumed total variance σ^2 (=1, in these cases). In addition, the retrieved ρ values based on (4) with the estimates of ζ^2 and η^2 show consistency with their counterparts of the simulation parameter.

In summary, with one group of subjects, LME provides a better choice for FPR control than SWP across the whole range of ρ values. On the other hand, even though SWP is virtually ideal in comparing two groups, LME offers a reasonably well-behaved alternative especially when there are 20 or more subjects in each group or when ρ is not too close to 0 (both of which are standard cases in MRI studies focusing on GM). Finally, the DF adjustments in the LME implementation appear appropriate.

Performance comparisons with experimental data

To demonstrate the performance of LME and to compare it with the best nonparametric methods from Part I for ISC analysis at the group level in real FMRI data, we utilize here the same experimental data from a naturalistic task FMRI session (Chen et al., 2016). Briefly, n = 48 healthy volunteers (n_1 = 24 males, n_2 = 24 females, age mean ± SD = 33.6 ± 5.7 and 34.7 ± 6.0 years old for males and females, respectively) watched six movie clips, each with an average length of two and half minutes, in a 3.0-T Siemens Trio scanner. Half of the six clips depicted mostly positive emotional episodes while the other half were of negative emotional valence. The series of clips were separated by a black screen for 10-30 s and preceded by a fixation cross for 30 s, leading to a total scanning time of 1,050 seconds. Scan parameters for the acquired whole brain BOLD EPI data were: voxel size of 3.8 × 3.8 × 4.0 mm^3, 36 axial slices, TR = 2,000 ms, TE = 30 ms, in plane FOV = 240 × 240 mm^2, flip angle = 90°.

The EPI time series went through the following preprocessing steps in AFNI (the precise *afni_proc.py* command which was used for processing is provided in Appendix B): de-spiking, slice timing and head motion corrections, affine alignment with anatomy, nonlinear alignment to a Talairach template (TT_N27) using 3dQwarp (all transformations were combined, per time point, to avoid repeated interpolation) at a voxel size of 3.5 × 3.5 × 3.5 mm^3, and smoothing with an isotropic FWHM of 6 mm. FreeSurfers *recon-all* command was used to estimate tissue maps for each subject, and the lateral ventricle and white matter (WM) maps were then eroded. Regressors were created from the first three principal components of the ventricles, and fast ANATICOR (Jo et al., 2010) was implemented to provide local WM regressors. Additionally, the subject's 6 motion time series, their derivatives and linear polynomial baselines for each of the 6 movie runs were included as regressors, for a total of 28. Censoring of time points was performed whenever the per-time motion (Eucliean norm of the motion derivatives) was > 0.3 mm or when ≥ 10% of the brain voxels were outliers.

ISC was computed over 406 time points (having excluded the periods of fixation and blank screen) at the voxel level between all pairs of the n = 48 subjects using 3dTcorrelate in AFNI, leading to N = 48 × 47/2 = 1,128 ISC values per voxel. The computation time for the SWB/SWP was approximately 0.5 hours for the case of one group (n = 24 subjects, with 24 CPUs) and 1.3 hours for two groups (n = 48 subjects, with 24 CPUs) on a Linux system (Fedora 14) with Intel textsuperscript® Xeon® X5650 at 2.67GHz. The runtime for 3dLME execution was about 15 mins for each of the one and two group cases (using 12 CPUs).

Of interest here were statistical inferences of the ISC for each group (i.e., one-sample tests for each of $\boldsymbol{R}_{11}$ and $\boldsymbol{R}_{22}$), the difference between the two sexes (direct comparison of WGC,

$\boldsymbol{R}_{11}$ vs $\boldsymbol{R}_{22}$), the BGC, and each WGC in contrast to the BGC (indirect comparisons of $\boldsymbol{R}_{11}$ vs $\boldsymbol{R}_{21}$ and $\boldsymbol{R}_{22}$ vs $\boldsymbol{R}_{21}$); that is, totally there six tests: two WGCs, one between-group comparison, one BGC, and two within- versus between-group, were performed using AFNI programs 3dLME. The results of the within-group (male) and between-group comparison are illustrated in Fig. 5 while the performance for BGC and the other two comparisons ($\boldsymbol{R_{11}}$ vs $\boldsymbol{R_{21}}$ and $\boldsymbol{R_{22}}$ vs $\boldsymbol{R_{21}}$) are shown in Fig. 8 of Appendix C.

For the male group, LME detected slightly more voxels than SWB under the voxel-wise significance level of 0.001 (the first two images in the upper panel, Fig. 5). Interestingly, those extra voxels from LME (yellow in the second image one the first row, Fig. 5) had small (below 0.1) group ISC values, and they corresponded to low (below 0.1) correlation parameter ρ values (green in the third image on the first row Fig. 5.). Nevertheless, most of those extra voxels detected by LME, despite their low ISC values, were robust and effective, considering the voxel-wise significance level of 0.001 and the large cluster sizes. We note that the whole brain is largely synchronized across the subjects. And we also emphasize that the higher detection power of LME than SWB is consistent with the simulation results (first and third rows in Fig. 3): even though both methods tend to be over-conservative when $\rho <$ 0.1, LME is less so than SWB.

On the other hand, for the comparison between males and females, LME and SWT achieved roughly the same level of detection (voxel-wise significance of 0.05, the first two images in the lower panel, Fig. 5). The tests involving the BGC are shown in Appendix C. More revealingly, the estimates of the cross- and within-subject variances, ζ^2 and η^2) can be directly estimated through LME (the last two columns in Fig. 5). The cross-subjects variance ζ^2 tended to roughly follow the same pattern of the ISC map (the first image in the upper panel, Fig. 5)), indicating that the regions that were highly synchronized across subjects had more heterogeneities than those that were slightly synchronized or unsynchronized. On the other hand, the within-subject (or residual) variance η^2 also had a similar spatial pattern as that of the group ISC map (i.e., large values in similar locations), but to a lesser extent than ζ^2. As expected, most voxels in WM and CSF had estimated η^2 values which were in the neighborhood of the approximate lower bound, η^2_{min}, derived in (5). In addition, voxels with statistically significant ISC values typically had estimated η^2 values that were much greater than the lower bound, except for a small proportion that had $\eta^2 \approx \eta^2_{min}$.

In part I, we had to guess the rough range of ρ values in the brain by comparing the detection power across various nonparametric methods relative to the simulation results. In contrast, the LME approach can directly estimate the correlation parameter ρ value at each voxel (the third column in Fig. 5), and, with the real experimental dataset, the estimated ρ values in the brain show an interestingly similar pattern to the ISC map (the first image in the upper panel, Fig. 5. Specifically, ρ is in the interval of [0.3, 0.5] in the highly synchronized regions, around 0.25 in the moderately synchronized regions, and below 0.1 in the rest of the brain. We note that the relative magnitude of group ISC values followed a similar pattern to ρ, with largest ISC in locations of highest ρ, etc.

Discussion

LME as an extension to GLM

In the conventional regression or GLM, the data are typically assumed to come from a similar, homogeneous sample (or samples), in the sense that the residuals are presumed to be identically and independently distributed. However, under some scenarios (e.g., the ISC paradigm) the acquired data are stratified with various subjects across which heterogeneities may occur, and a more adequate model would include and account for subject-specific adjustments (or "random intercept" in the LME terminology). It is due to the presence of this across-subjects heterogeneity that we prefer to use the subject-specific presentation of the LME-CRE formulation (2) or (9) for parametric modeling, as opposed to using *t*-tests or GLMs. Simply ignoring the stratification or clustering would lead to distorted statistical inferences, as shown in simulations and analytical results from experimental datasets in Part I, as well as in the methods adopted in the literature. Unlike the single source of variance (residuals) in regression or GLM, two or more sources of variances are considered in LME. In this perspective, the $\rho = 0$ case reduces the LME to the conventional GLM by removing the two random effects, and only in this special circumstance can the Student's *t*-test or EW nonparametric approaches that have been adopted in the literature can be justified.

As an extension of GLM, LME has the great flexibility to account for various variance-covariance structures among clustered data, and its modeling techniques still undergo constant development in modeling and numerical implementation. The adjective *mixed* in LME reflects the fact that both fixed and random effects are considered and simultaneously estimated within the same model. The framework can be viewed as a combination of or compromise between Bayesian and frequentist approaches: an LME model is constructed through a hierarchical (or conditional) structure where the parameters are considered as random variables (Bayesian perspective); however, those parameters are not arbitrarily specified a priori as hyperparameters, as in the Bayesian domain, but they are estimated from the data itself (frequentist domain) (for more discussion see, for example, Demidenko, 2013).

Adoption of the LME/CRE approach to ISC data analysis at the group level

By adopting the LME framework here, we decompose the individual ISC values into multiple constituent components and develop an analytical scheme that is adaptive to various scenarios. Relative to group (fixed) effects, each ISC value is conceptualized as having two independent adjustments (or random intercepts, in the LME terminology), associated with the respective subjects. The fixed effects obtained through REML reflect the best estimation of the model formulation, and they can be extendable to those potential subjects that were not recruited for the study on hand. As each individual subject may have a higher or lower deviation from the group effects, the random effects serve the purpose of tuning the best calibration under the model. Even though CRE can be subsumed into the standard LME formalism (8), the explicit platform (9) is more revealing and convenient in terms of the presentation and implementation, thanks to its distinctive independence property among the random components.

Specifically, as each ISC observation z_{ij} is associated with the combination of subjects i and j, we adopt the CRE framework (6), with double random intercepts and with both lower and upper triangular subsets in the ISC matrix $\boldsymbol{Z}$ as input data for the response variable. For the case with one group, the data doubling step means that the whole ISC matrix (except the diagonals) are incorporated in the LME model. With two groups, whenever $\boldsymbol{Z}_{21}$ or the lower triangular subset of $\boldsymbol{Z}_{11}/\boldsymbol{Z}_{22}$ is involved, their counterpart $\boldsymbol{Z}_{21}$ or the upper triangular subset of $\boldsymbol{Z}_{11}/\boldsymbol{Z}_{22}$ is also utilized through data doubling.

The tricky aspect for the ISC data in the LME framework (9) is the presence of double random intercepts: even though the two random effects are essentially just the same (one being recycled due to the symmetric nature of correlation coefficient), the index restraint of $i > j$ in the LME model (9) creates a scenario where the two random effects are not fully crossed. By loosening the constraint, we create an LME system having fully crossed random effects as well as balanced structures for both of the random-effects model matrices, leading to an identical variance estimate for each. To discount for the double usage of the data, we adjust the standard error out of the REML estimates as well as the DFs for the statistics, approaches which the consistency of the simulation results verified as being acceptable. By comparing to previous results in Part I, our LME modeling strategy has been validated through both simulations in FPR and power assessment as well as results from an experimental dataset.

The LME/CRE formulation not only allows us to properly partition the effects, but also offers an important feature: the variance-covariance structure that is naturally embedded in the model through two crossed random effects, θ_i and θ_j, can be explicitly estimated from the data. Unlike in conventional parametric models, where differentiations are explicitly made among paired, one- or two-sample t-tests, and ANOVA (or similarly for nonparametric methods such as permutations or bootstrapping), a single, comprehensive framework for LME is adaptable to various scenarios with between- and within-subject explanatory variables (Baaven et al., 2008). In practice, through a data duplication procedure, we achieve a balanced data structure with fully crossed random effect components, leading to easy implementation with the standard REML algorithm for LME as well as to a consistent estimation of the variances of crossed random effects (e.g., $\zeta_\pi^2 = \zeta_\lambda^2$ in (9)). To counter the effect of data redundancy, we propose that the resulting inflation in statistical inferences be compensated by adjustments in the number of degrees of freedom. The LME methodology with crossed random effects (9) will be soon added into the publicly available AFNI (Cox, 1996) program 3dLME (Chen et al., 2013), which was originally developed for the general LME platform (8).

Adaptability and advantages of LME/ CRE

In the same manner that the standard LME system (8) has a high adaptability, so the formulation (6) for ISC data can be easily extended to two or more groups and to include other between-subject variables, such as quantitative covariates, through augmenting the fixed-effects model matrix $\boldsymbol{X}$. For example, any quantitative covariates (e.g., age or IQ) and subjects from m groups can be modeled through adding $m - 1$ dummy-coded columns to $\boldsymbol{X}$. Furthermore, the LME approach to ISC data can be naturally adapted to include within-

subject factors (e.g., different types of viewed scenes), which cannot be easily handled under the standard parametric framework nor readily within the nonparametric frameworks discussed in Part I. The ability to suit a wide array of variable types and model organization is a powerful feature of the LME/CRE framework.

The analytical depth of LME offers multiple advantages over the conventional ANOVA or GLM due to its broader flexibility to deal with data that are acquired from related (or clustered) measuring units. 1) One generic and uniform LME platform can be constructed to incorporate and adapt to both categorical and quantitative variables. 2) LME is capable of handling missing data as long as they occur at random. 3) LME provides various systematic and principled approaches to characterizing heteroscedasticity and non-spherical (or aspherical) variance-covariance structure of the random effect components (Bates et al., 2015). For example, the cross-subject variance ζ^2 was assumed to be the same between the two groups in our analysis with the experimental data (fourth column on the second row, Fig. 5), the comparison with the case of one group (fourth column on the first row, Fig. 5) justifies the homoscedasticity assumption. However, it is convenient to specify different cross-subject variances across groups, if desired, within the LME framework.

4) LME provides a platform by which all desired explanatory variables can be incorporated simultaneously into one model. Due to practical considerations, such as computational power and heterogeneous stimulus timing across subjects, the current practice with FMRI data analysis pipeline inherits the historical approach of splitting the whole analysis into two major steps, individual and group levels, with the assumption that the effect estimates from the individual level are equally reliable. A better approach would be to adopt LME by combining all the subjects' data into a hierarchical model, while an alternative is to take both the effect estimates and their reliability information together to the group analysis (Worsley et al., 2002; Woolrich et al., 2004; Chen et al., 2012).

5) Ideally, model building should be a crucial process in statistical inferences, and LME provides an irreplaceable tool in that regard. For example, a simple model, while easy to analyze, may tend to strongly distort what the real data reveal. On the other hand, an overly complicated model through over-parameterization could lead to high cost in the number of degrees of freedom, statistical power, numerical instability, and even an underdetermined system. In addition to the enhanced statistical power, LME allows the investigator to apply parsimonious principles or Occam's razor, and to achieve a balanced compromise between model complexity and faithfulness to the data (Baaven et al., 2008). Typically characterized by criterion indices such as AIC and BIC, the balance is directly reflected in the two components of the profiled deviance function[6] upon which the optimization is achieved through combining the fidelity of the fitted data to the observed data and model complexity. The over-parameterization also explains why LME is conservative when the correlation parameter ρ is close to 0 (green curve on the first row, Fig. 3): power loss follows when one overfits the data with random effects (or correlation structure) that are nonexistent or negligible. Here, the same phenomenon occurs in those regions in the brain where ρ is close to 0 (Fig. 5), since that implies full independence and would not require modeling with two

[6]The profiled deviation is negative twice the log-likelihood function of the LME model.

random effects (in effect, such modeling then technically represents an 'over-parameterization'). However, in practice when using FMRI, this tends to occur in regions which are not of interest (such as WM and CSF), and the impact itself is quite small, particularly with standard group sizes (e.g., 20 or more). Therefore, adjustments for these few cases are not needed within the massively univariate FMRI approach.

Some of these advantages also apply to the comparison between LME and the nonparametric approaches explored in Part I, in particular in providing more powerful detections (e.g., statistically more sensitive than SWB when $\rho < 0.1$) and interpretations as well as more detailed characterization of the data. For example, by using the standard Gaussianity assumption, the insight about the effects for each individual subject is directly characterized by the deviation θ_i and the cross-subject variance component ζ^2 in (2). Furthermore, the correlation estimate ρ and residual or within-subject variance η^2 are also part of the output. As shown in Fig. 5, the heterogeneity of the estimated ρ values across the brain may reveal invalid assumptions, outlying subjects,or processing issues (e.g., suboptimal alignment, superfluous spatial smoothing). Another advantage of LME is its higher detection power with one group of subjects than SWB especially when the correlation parameter ρ is low. Lastly, in addition to the various flexibilities discussed above, such as the capability of handling missing data and various variance-covariance structures, LME is computationally more economical and advantageous even compared to the implementations that the nonparametric approaches can handle. For instance, each scenario of analyzing one or two groups would have to be implemented with a different nonparametric method and interface. In contrast, the design matrix $\boldsymbol{X}$ in the formulation (9) plays a role of placeholder for fixed effects and allows the flexibility to model unlimited number of explanatory variables such as between- or within-subject factors as well as quantitative (between- or within-subject) covariates or confounding effects.

Robustness and limitations of LME/CRE

It is of note that the interpretation power and model/data quality check come at some cost as well. LME requires stronger assumptions (linearity and Gaussianity) than independence and exchangeability in nonparametric methods. If the assumptions are strongly violated (e.g., skewed data, outliers), higher false positive or false negative rates may occur. In addition, the correlation structure between the two crossed random effects, embodied by the parameter ρ, is assumed even though in some brain regions (e.g., white matter) this may be an excessive parameterization (see discussion in the previous section). The impact of over-parameterization is shown though our simulations with a small over-conservative performance in controlling for FPR when $\rho = 0$ (first two rows in Fig. 3). In contrast, the nonparametric SWB approach is even worse in controlling FPR when $\rho = 0$ with one group, while SWP performs virtually ideal with two groups.

More specifically, the LME estimation for the fixed effects would still be unbiased even if the random effects are inaccurately specified or when the Gaussianity or homoscedascity assumption is violated (Pinheiro and Bates, 2004). Nevertheless, the misspecification or distributional violation usually leads to potentially misaligned statistical inferences due to inaccurate estimate for the standard error of each fixed effect (Verbeke and Molenberghs,

2000). Under those circumstances, a nonparametric approach, if feasible (e.g., cases with one or two groups without covariates), might be a better choice. However, no clear-cut recommendation can be offered regarding method decision because of the complexity with different modeling capabilities as well as the difficulty in model building, tuning or comparing when simultaneously analyzing a large number of voxels in the brain.

Interpreting ρ as intraclass correlation (ICC)

When applied to the ISC data, the two random effect components in the LME model (9) share the same variance, and lead to the multiplicative factor of 2 in the correlation coefficient (4). Interestingly, we note that the ρ value as a variance ratio can be recognized as the direct definition of intraclass correlation (ICC) here in an LME context. In contrast to the upper bound shown through an abstract proof with eigenvalues in Part I, this ICC interpretation under the LME framework gives a direct and meaningful explanation as to why $\rho \leq 0.5$: there are two random intercepts in the LME system (2), i.e., two copies of the cross-subject variance ζ^2 in the variance partitioning.

The ICC in general can be interpreted in four perspectives. 1) The ICC is the proportion of total variance that is attributed to a random factor (or accounted for by the association across the levels of the random factor); that is, as the variance associated with the random factor increases, it is assumed to be less likely that the levels (e.g., subjects in the ISC context) within the factor are similar. 2) The ICC is the expected correlation between two responses that are randomly drawn among the factor levels (e.g., the correlation of two ISC values, z_{ij} and z_{jk}, associated with the same subject j, which is exactly the original definition of ρ in **Introduction**). When some subjects have generally higher (or lower) ISC values relative to the group (or fixed) effects, or when there is some extent of consistency among all the ISC values associated with a specific subject, then those ISC values are correlated, and the formulation (4) basically captures that correlation or consistency. 3) ICC is an indicator of homogeneity of the responses among the levels (e.g., subjects) of the random factor: a higher ICC means more similar or homogeneous responses. 4) It shows the extent of common environments that responses share. The ICC would be larger if responses associated with the same factor level (e.g., subject) are under more similar environments and have closer values.

Within the ISC context, the ICC interpretations of ρ may even be useful in model quality check. For instance, when ρ is close to zero, the ISC values associated with a subject are no more similar than those from different subjects, and the random effect components could be removed to avoid unnecessary over-parameterization. In addition, the spatial heterogeneities of the ICC ρ across the brain, as well as the between- and within-subject variances ζ^2 and η^2, are revealing as shown in the results from the real data analysis (Fig. 5 and Fig. 8).

Relation of ISC to multivariate distance matrix regression (MDMR)

With the definition of the distance between a pair of subjects i and j as $\sqrt{2(1-r_{ij})}$ (Gower and Krzanowski, 1999), where r_{ij} is the ISC value at a voxel, the ISC data may benefit from combination with a multivariate distance matrix regression (MDMR) approach. MDMR allows for between-subjects explanatory variables and has been adopted in FMRI for investigating resting-state data (e.g., Shehzad et al., 2014). When the Gaussianity

assumption of the ISC data is severely violated, MDMR may well be a better choice, thanks to the permutation testing employed in evaluating the significance level of an effect for the associated pseudo-*F* statistic. However, we also note that in most practical situations in FMRI, the presented LME approach is advantageous to MDMR in the following aspects: 1) LME possesses greater modeling capability (e.g., allowing for within- subject variables); 2) unlike in LME, each effect in MDMR is assessed by the associated pseudo-*F* statistic, but the effect estimate is not available, nor is its directionality (positive or negative) known.

Conclusion

The LME modeling provides a well-suited platform for the ISC data at the group level with each ISC value partitioned into fixed effects and two random intercepts. Through a data duplication step, we construct testing statistics that achieve proper FPR, without resorting to the computationally expensive approaches in statistical inferences. In addition, the LME flexibility allows the investigator to conveniently include various fixed effects including within-subject explanatory variables. Other benefits are strong interpretation power, relatively low computational cost, and data/model quality control. The LME modeling strategy and statistical inference scheme will be added into the AFNI program 3dLME that was originally developed for neuroimaging group analysis with a general LME interface (Chen et al., 2013). In addition to omnibus *F*-statistics for main effects and interactions as well as fixed effect estimates and their *t*-statistics, 3dLME also allows various post hoc tests.

Acknowledgments

Our work benefited significantly from the statistical computational language and environment *R*, the LME package lme4 with the capability to modeling crossed random effects, and the great support of the *R* community. The research and writing of the paper were supported by the NIMH and NINDS Intramural Research Programs (ZICMH002888) of the NIH/HHS, USA, and by a grant (2012-0230) from the Asan Institute for Life Sciences, Asan Medical Center, Seoul, Korea.

Appendix A. List of acronyms used in the paper

ANOVA	**analysis of variances**	**GLM**	**general linear model**
BLUE	best linear unbiased estimate	ICC	intraclass correlation
BLUP	best linear unbiased predictor	ISC	inter-subject correlation
CRE	(LME with) crossed random effects	LME	linear mixed-effects
DF	degrees of freedom	REML	restricted maximum likelihood
EW	element-wise	SW	subject-wise
FPR	false positive rate		

Appendix B. FMRI processing

The general sequence of FMRI data preprocessing steps was described in the Methods section of this paper. However, for greater specificity and reproducibility, in this section we also include the exact command that was implemented for the processing. While there were several processing steps (or blocks) specified, each with many user-chosen options, it is possible to provide the exact pipeline in a succinct manner because the processing blocks

and options were created and specified using *afni_proc.py* in AFNI (v16.1.16; Cox, 1996). This tool permits the user full freedom to tailor a desired pipeline that may be reliably duplicated for the entire group, stored for future reference and published with a study for unambiguous description.

In this study FreeSurfer's *recon-all* was first run on the anatomical volume in order to produce tissue segmentation maps used in the processing (v5.3.0; Fischl et al.). The FreeSurfer output was converted into NIFTI format for use in AFNI using SUM.Vs *@SUMA_MakeSpec_FS* command (Saad et al., 2004; Saad and Reynolds, 2012), and then maps of the WM and ventricles were selected. The commands for these steps are provided in Table 1. In Table 2 we include the *afni_proc.py* command used in the present study (which is about 25 lines; based on the help file's Example 11) that generates a full, executable processing pipeline of > 500 lines.

Appendix C. ISC analysis results with experimental dataset involving BGC

The group analyses involving the BGC $\boldsymbol{Z}_{21}$ are shown here in Figures 6, 7, and 8, as counterparts of Figures 3, 4, and 5.

References

Baayen R, Davidson D, Bates D. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.

Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015; 67(1):l–48. DOI: 10.18637/jss.v067.i01

Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR. and for the Alzheimer's Disease Neuroimaging Initiative. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. Neuroimage. 2013; 66:249–260. [PubMed: 23123680]

Chen G, Saad ZS, Nath AR, Beauchamp MS, Cox RW. FMRI Group Analysis Combining Effect Estimates and Their Variances. NeuroImage. 2012; 60:747–765. [PubMed: 22245637]

Chen G, Saad ZS, Britton JC, Pine DS, Cox RW. Linear mixed-effects modeling approach to FMRI group analysis. NeuroImage. 2013; 73:176–190. [PubMed: 23376789]

Chen, G., Shin, Y-W., Taylor, PA., Glen, D., Reynolds, RC., Israel, RB., Cox, RW. Untangling the Relatedness among Correlations, Part I: Nonparametric Approaches to Inter-Subject Correlation Analysis at the Group Level. NeuroImage. 2016. (in press), http://dx.doi.org/10.1016/j.neuroimage.2016.05.023

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 1996; 29:162–173. http://afni.nimh.nih.gov/afni. [PubMed: 8812068]

Demidenko, E. Mixed Models: Theory and Applications with R. Wiley; 2013.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33(3):341–55. [PubMed: 11832223]

Gower JC, Krzanowski WJ. Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. Journal of the Royal Statistical Society: Series C. 1999; 48:505–519.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. Intersubject synchronization of cortical activity during natural vision. Science. 2004; 303:1634–1640. [PubMed: 15016991]

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. A hierarchy of temporal receptive windows in human cortex. Journal of Neuroscience. 2008; 28(10):2539–2550. [PubMed: 18322098]

Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW. Mapping sources of correlation in resting state FMRI, with artifact detection and removal. Neuroimage. 2010; 52(2):571–582. [PubMed: 20420926]

Kauppi J-P, Pajula J, Tohka J. A Versatile Software Package for Inter-subject Correlation Based Analyses of fMRI. Frontiers in Neuroinformatics. 2014; 8:2. [PubMed: 24550818]

Pinheiro, JC., Bates, DM. Mixed-Effects Models in S and S-PLUS. Springer; 2004.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2016. URL https://www.R-project.org/

Saad, ZS., Reynolds, RC., Argall, BD., Japee, S., Cox, RW. IEEE International Symposium on Biomedical Imaging: from Nano to Macro. IEEE; Arlington VA: 2004. SUMA: An interface for surface-based intra- and inter-subject analysis within AFNI; p. 1510-1513.

Saad ZS, Reynolds RC. SUMA. Neuroimage. 2012; 62:768–773. [PubMed: 21945692]

Shehzad Z, Kelly C, Reiss PT, Craddock RC, Emerson JW, McMahon K, Copland DA, Castellanos FX, Milham MP. An Multivariate Distance-Based Analytic Framework for Conneetome- Wide Association Studies. NeuroImage. 2014; 93(1):74–94. [PubMed: 24583255]

Sheskin, DJ. Parametric and Nonparametric Statistical Procedures. 3. Chapman & Hall/CRC; 2004.

Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. Springer Series in Statistics. 2000

Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM. Multilevel linear modelling for FMRI group analysis using Bayesian inference. NeuroImage. 2004; 21(4):1732–1747. [PubMed: 15050594]

Worsley KJ, Liao C, Aston J, Petre V, Duncan GH, Morales F, Evans AC. A general statistical analysis for fMRI data. Neuroimage. 2002; 15:1–15. [PubMed: 11771969]

$$\boldsymbol{R}^{(n)} = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_n \end{array} \begin{array}{c} \begin{array}{ccccc} S_1 & S_2 & S_3 & \cdots & S_n \end{array} \\ \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & 1 & r_{23} & \cdots & r_{2n} \\ r_{31} & r_{32} & 1 & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & 1 \end{pmatrix} \end{array} \qquad \boldsymbol{Z}^{(n)} = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_n \end{array} \begin{array}{c} \begin{array}{ccccc} S_1 & S_2 & S_3 & \cdots & S_n \end{array} \\ \begin{pmatrix} - & z_{12} & z_{13} & \cdots & z_{1n} \\ z_{21} & - & z_{23} & \cdots & z_{2n} \\ z_{31} & z_{32} & - & \cdots & z_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & z_{n3} & \cdots & - \end{pmatrix} \end{array}$$

Figure 1.

ISC matrix $\boldsymbol{R}^{(n)}$ and its Fisher-transformed counterpart $\boldsymbol{Z}^{(n)}$ with one group of n subjects. Due to the symmetry, only half of the off-diagonal elements (shaded in gray) are usually considered in group analysis.

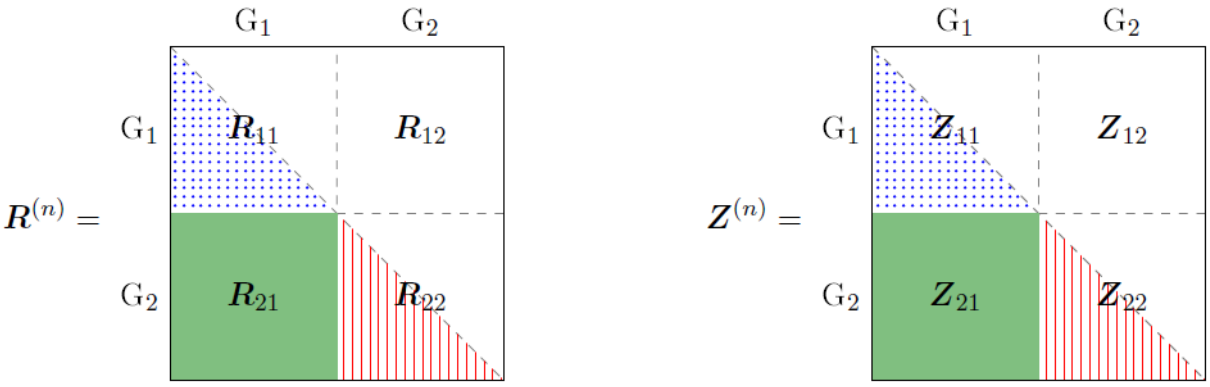

Figure 2.

Schematic illustration of $\boldsymbol{R}^{(n)}$ and the Fisher-transformed ISC data $\boldsymbol{Z}^{(n)}$ for two groups, G_1 and G_2, highlighting the lower triangular elements. With n_1 and n_2 subjects, respectively, in the two groups G_1 and G_2, the $N_1=\frac{1}{2}n_1(n_1-1)$ elements in $\boldsymbol{Z}_{11}$ (blue) and $N_2=\frac{1}{2}n_2(n_2-1)$ elements in $\boldsymbol{Z}_{22}$ (red) are WGC values while the $N_{12} = n_1n_2$ elements in Z_{21} (green) show the BGC values. Six meaningful group tests can be formulated: three effects $\boldsymbol{Z}_{11}$, $\boldsymbol{Z}_{21}$, and $\boldsymbol{Z}_{22}$, and three comparisons $\boldsymbol{Z}_{11}$ vs $\boldsymbol{Z}_{22}$, $\boldsymbol{Z}_{11}$ vs $\boldsymbol{Z}_{21}$, and $\boldsymbol{Z}_{22}$ vs $\boldsymbol{Z}_{21}$, as discussed in the text.

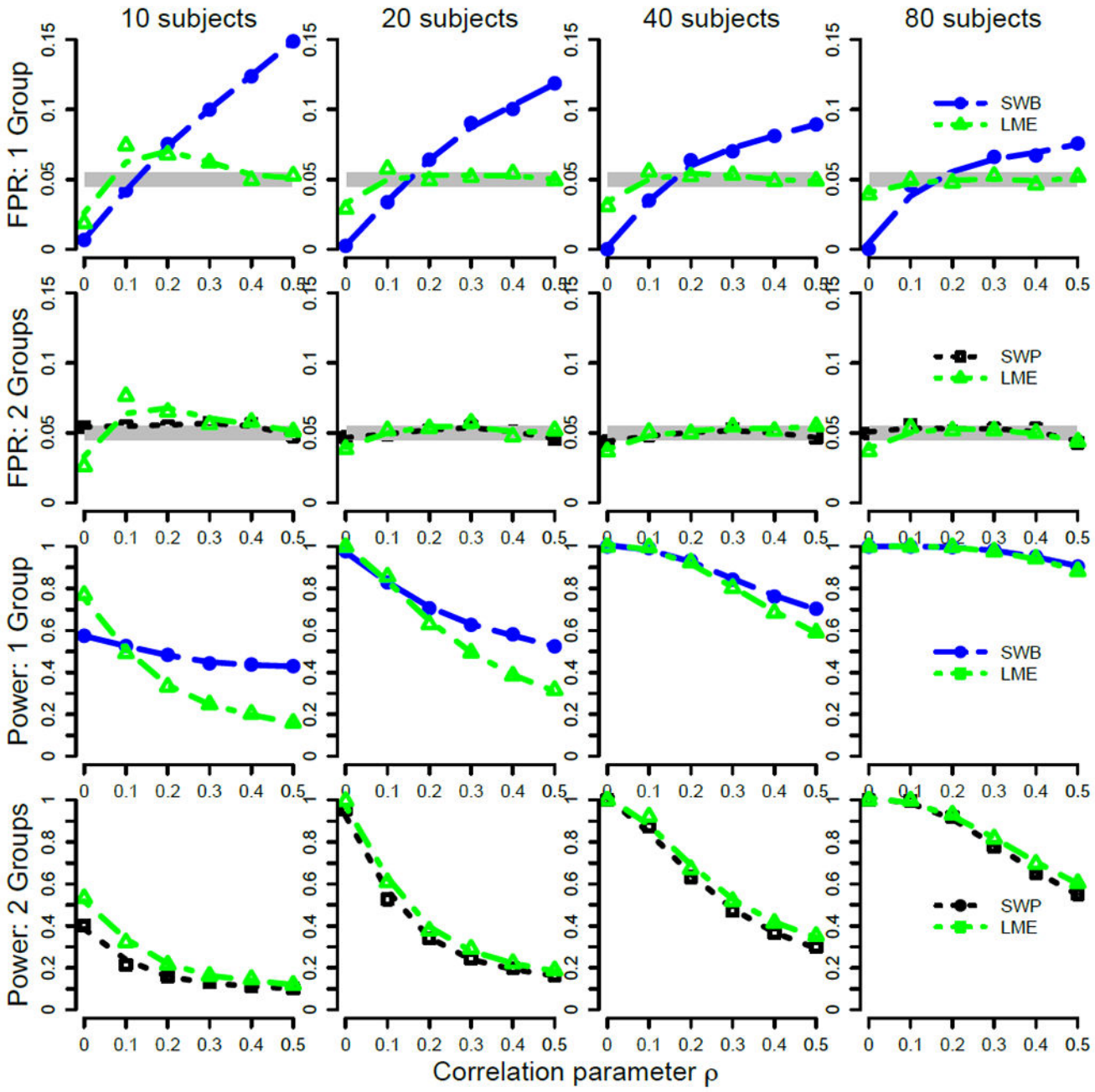

Figure 3.

Simulation parameters and results are shown here for three methods: SWB, SWP and LME. False positive rate (FPR) performances are illustrated in the first two rows, and power achievement in the last two rows. Each of the four columns represents the number of subjects in each group (one group, $n = 10, 20, 40, 80$; two groups, $n_1 = n_2 = 10, 20, 40, 80$). The gray band of FPR = 0.05 in the first two rows indicates the 95% confidence interval of the target (or nominal) value (with a width of ±0.012 for each simulation with 5000 realizations†). The curves for FPR and power were fitted to the simulation results (plotting symbols) through a cubic smoothing spline. Among the three possible comparisons for the

two-group scenario, only the direct contrast between the two WGCs, $\boldsymbol{Z}_{11}$ vs $\boldsymbol{Z}_{22}$, is shown here, but the results for other two indirect contrasts ($\boldsymbol{Z}_{11}$ vs $\boldsymbol{Z}_{21}$ and $\boldsymbol{Z}_{11}$ vs $\boldsymbol{Z}_{21}$) with SWP were similar (see Fig. 8 in Appendix C). The SWP testing is uniformly well-behaved and essentially ideal for two groups (black in the second row). On the other hand, in the one-group case, LME offers a uniformly better FPR control, even though it can be a little over-conservative (blue in the first row) when $\rho \leq 0.1$. Notice that the y-axis range of FPR, [0, 0.15] here (upper two panels), is different from that in Fig. 2 in Part I [0, 0.8], due to method variability.

†The confidence band is computed with the assumption of a binomial distribution B(n,p), where $n = 5000$, p = PFR = 0.05.

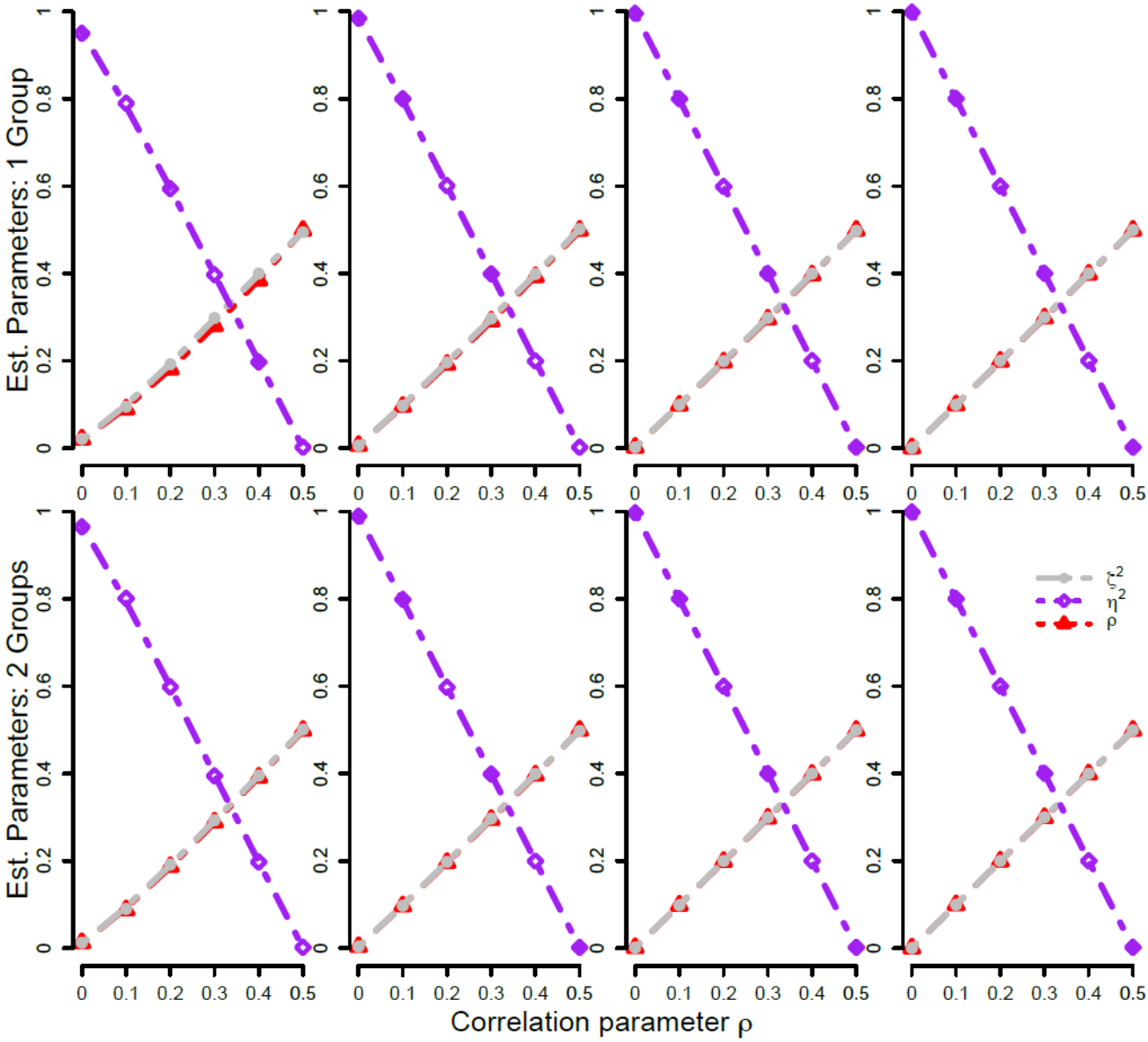

Figure 4.

Three parameters, correlation value ρ (red), cross- and within-subject variances ζ^2 (gray) and η^2 (purple) are estimated through LME with one and two groups, and then averaged across the 5000 simulations. The retrieved ρ values match reasonably well the six simulation parameter values of ρ, and the estimated variances ζ^2 and η^2 satisfy the variance decomposition $2\zeta^2 + \eta^2 = \sigma^2$ (=1, here).

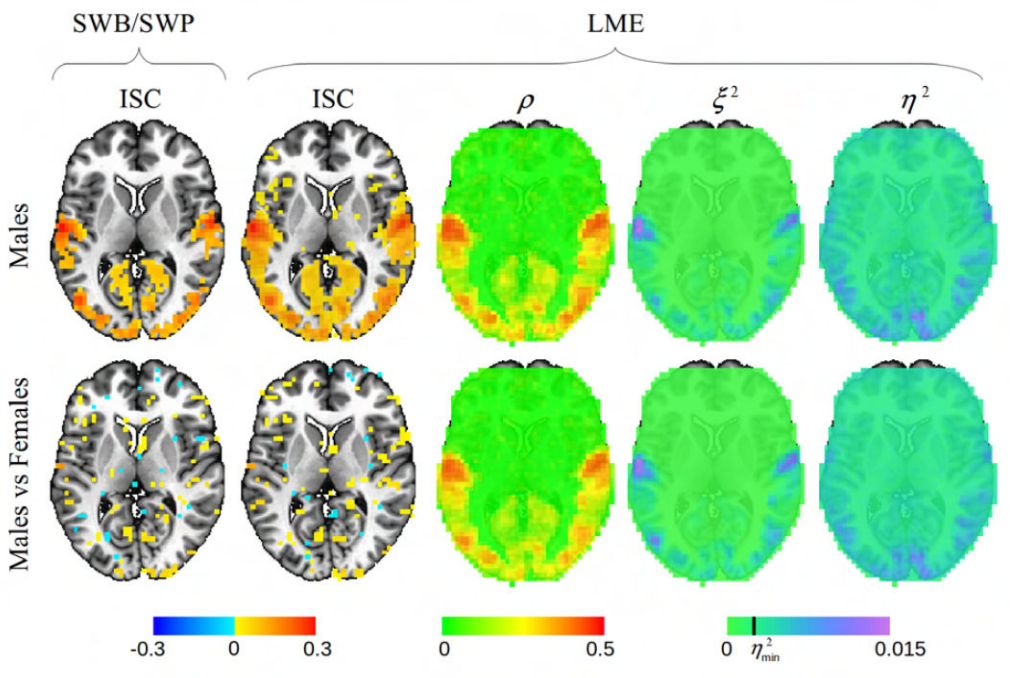

Figure 5.

Performance comparisons with an experimental dataset. Axial views ($Z = 8$ mm; radiological convention: subject left is right) of ISC group results (thresholded by p-values, below) of an experimental dataset are illustrated for three methods: SWB for males, SWP for group comparison, and LME. The colors code for ISC values in the first two columns (which were inverse Fisher-transformed from the z-values of the LME output), for the estimated ρ in the third column, and for the estimated between- and within-subject variances ξ^2 and η^2, respectively, in the last two columns. The colorization for the two variances is different from the first three columns due to the range differences, and the value of η^2_{min} in the colorbar, 0.00248, is an approximate lower bound for η^2 for time series with $T = 406$ points, as shown in (5). For the male group ($n = 24$, upper panel, two-tailed significance level $p = 0.001$) LME was slightly more powerful than SWB in small portions of the brain. For two-group comparison ($n_1 = n_2 = 24$, lower panel, two-tailed significance level $p = 0.05$), LME and SWP rendered very similar identifications. Their performances for the BGC, $\boldsymbol{R}_{11}$, and the other two indirect contrasts ($\boldsymbol{R}_{11}$ vs $\boldsymbol{R}_{21}$ and $\boldsymbol{R}_{22}$ vs $\boldsymbol{R}_{21}$) are shown in Fig. 8 of Appendix C. We note that 1) the parameters ρ, ξ^2, and η^2 for the group comparison (last three images in the lower panel) were estimated with the assumption of same variances across the groups (homoscedasticity); 2)multiple testing correction was not performed so that voxel-wise comparisons among the methods could be directly visualized; and 3) the three methods rendered virtually the same group estimate for ISC, but differed slightly in significance detection (i.e., the color at each voxel is roughly the same across the three testing methods if the significance survives the corresponding threshold).

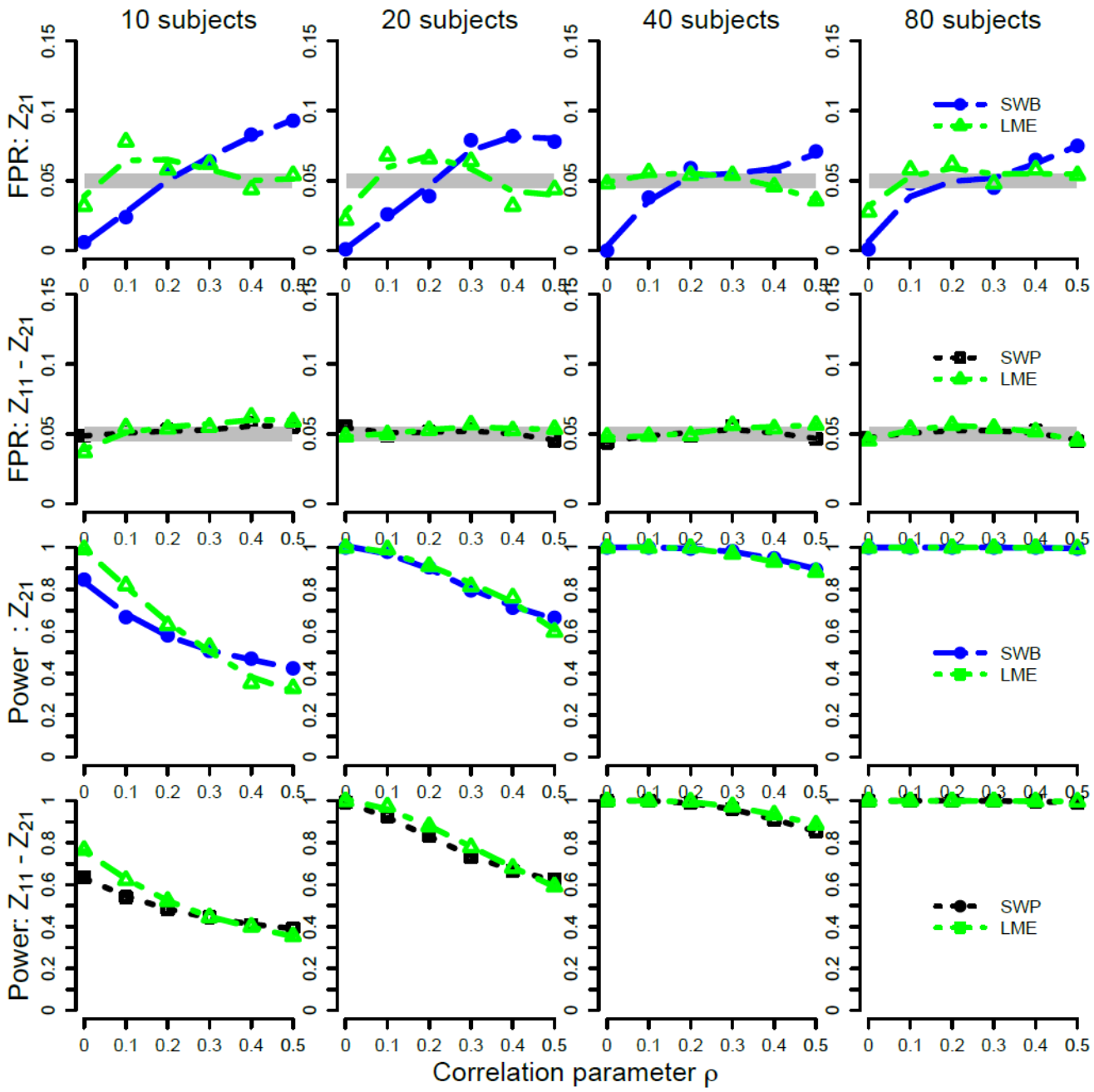

Figure 6.

Simulation parameters and results are shown here for $\boldsymbol{Z}_{21}$ and the contrast between $\boldsymbol{Z}_{11}$ and $\boldsymbol{Z}_{21}$ with three methods: SWB, SWP and LME. False positive rate (FPR) performances are illustrated in the first two rows, and power achievement in the last two rows. Each of the four columns represents the number of subjects in each group (one group, $n = 10, 20, 40, 80$; two groups, $n_1 = n_2 = 10, 20, 40, 80$). The gray band of FPR = 0.05 in the first two rows indicates the 95% confidence interval of the target (or nominal) value (with a width of ±0.012 for each simulation with 5000 realizations[†]). The curves for FPR and power were fitted to the simulation results (plotting symbols) through a cubic smoothing spline. The

SWP testing is uniformly well-behaved and essentially ideal for two groups (black in the second row). On the other hand, in the case with $\boldsymbol{Z}_{21}$, LME offers a uniformly better FPR control, even though it can be a little over-conservative (green in the first row) when $\rho \leq 0.1$. Notice that the y-axis range of FPR, [0, 0.15] here (upper two panels), same as in Fig. 3, is different from that in Fig. 2 in Part I [0, 0.8], due to method variability. The results for the contrast of $\boldsymbol{Z}_{22}$ and $\boldsymbol{Z}_{21}$ are virtually the same as that of $\boldsymbol{Z}_{11}$ and $\boldsymbol{Z}_{21}$ due to the symmetry.
† Tlie confidence band is computed with the assumption of a binomial distribution B(n, p), where $n = 5000$, p = PFR = 0.05.

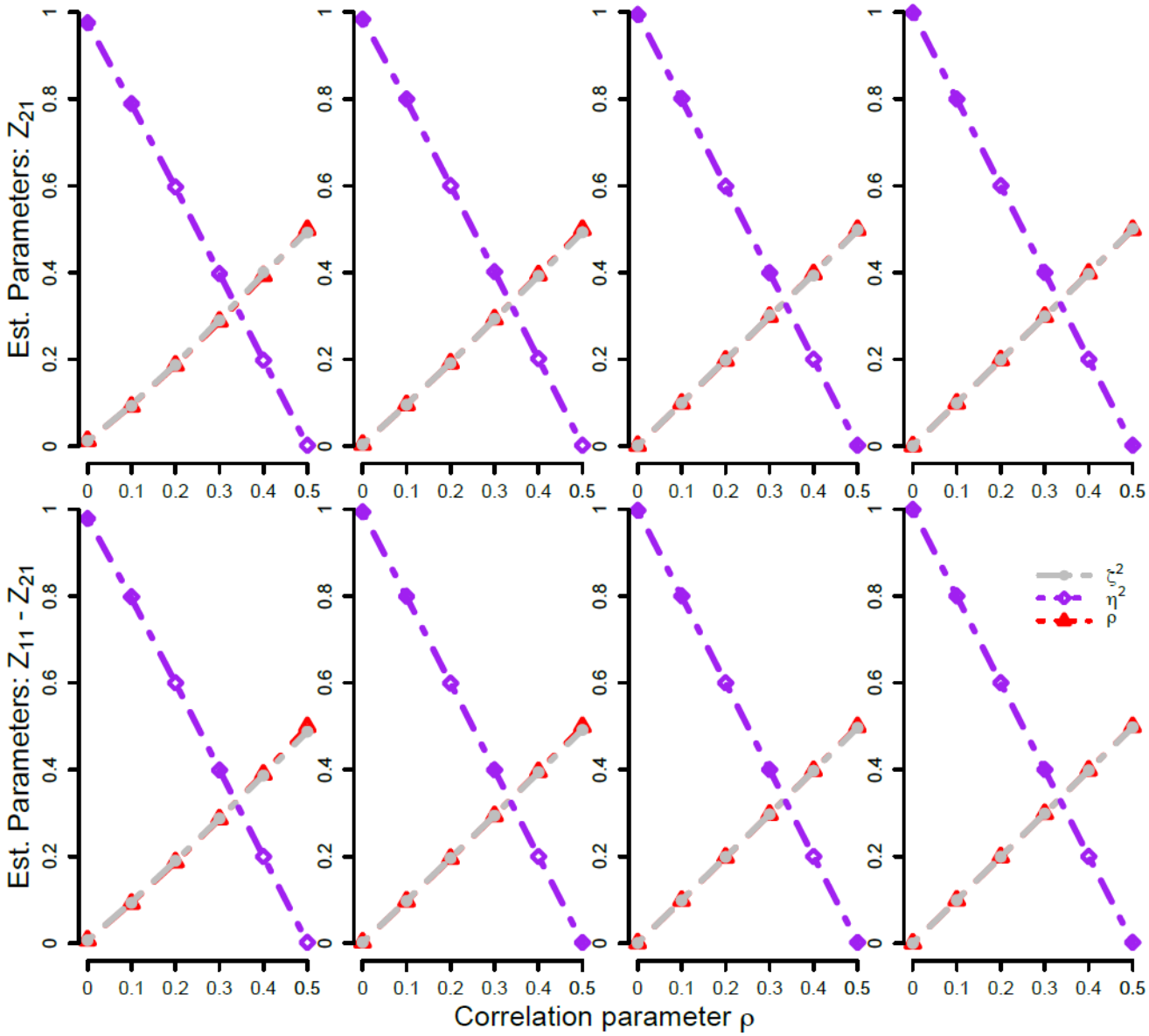

Figure 7.

Three parameters, correlation value ρ (red), cross- and within-subject variances ζ^2 ((gray) and η^2(purple) shown here are estimated through LME with the BGC Z_{21} and the contrast between Z_{11} and Z_{21}, and then averaged across the 5000 simulations. The retrieved ρ values match reasonably well the six simulation parameter values of ρ, and the estimated variances ζ^2 and η^2 satisfy the variance decomposition $2\zeta^2 + \eta^2 = \sigma^2 = 1$. The results for the contrast of Z_{22} and Z_{21} are virtually the same as that of Z_{11} and Z_{21} due to the symmetry.

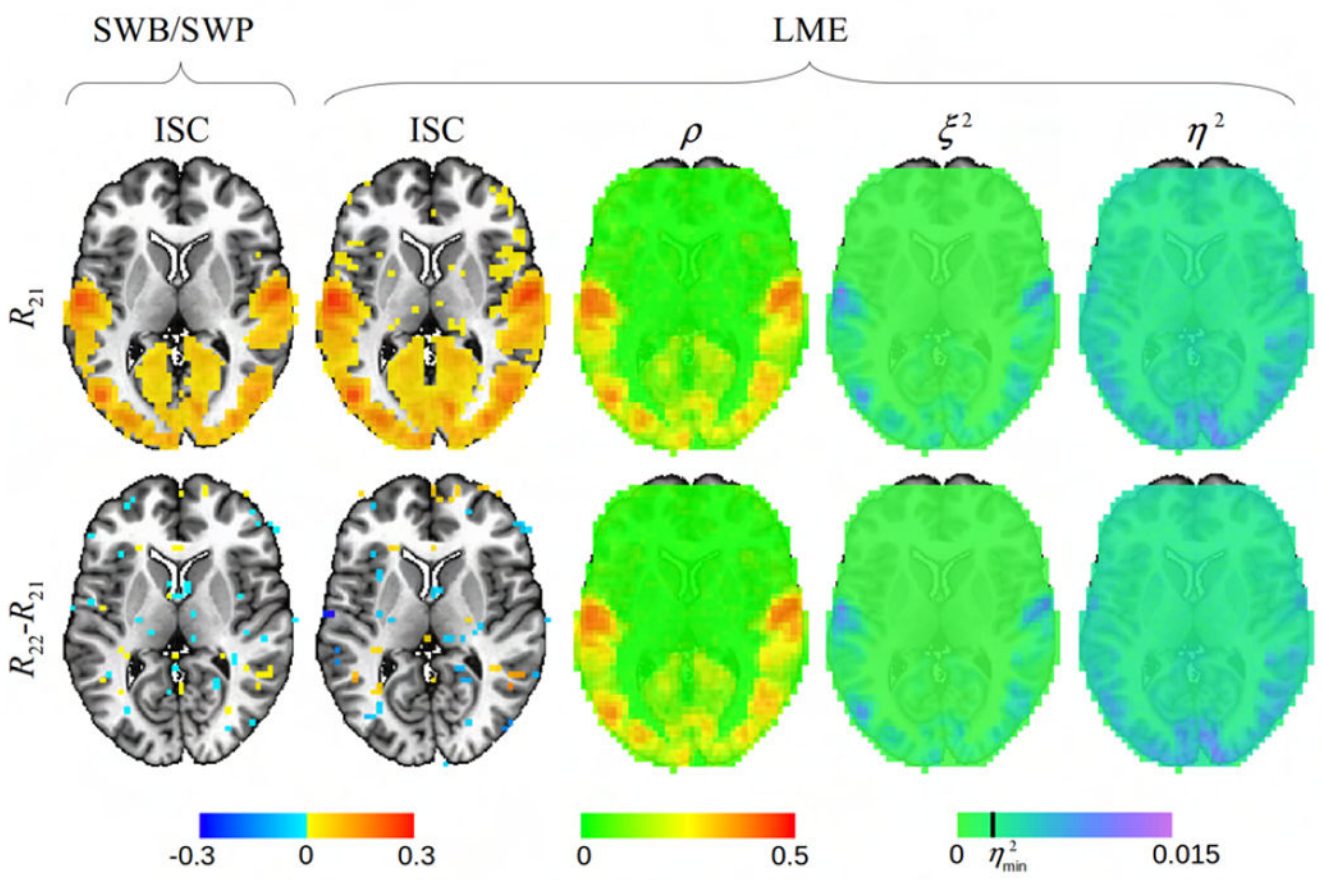

Figure 8.

Performance comparisons with an experimental dataset for the BGC, $\boldsymbol{R}_{21}$ and the contrast of $\boldsymbol{R}_{22}$ vs $\boldsymbol{R}_{21}$. Axial views (Z = 8 mm; radiological convention: left is right) of ISC group results (thresholded by p-values, below) of an experimental dataset are illustrated for three methods: SWB for $\boldsymbol{R}_{21}$, SWT for $\boldsymbol{R}_{22}$ vs $\boldsymbol{R}_{21}$, and LME. The contrast of $\boldsymbol{R}_{22}$ vs $\boldsymbol{R}_{21}$can be performed, but is not shown here. The colors code for ISC values in the first two columns (which were inverse Fisher-transformed from the z-values of the LME output), for the estimated ρ in the third column, and for the estimated between- and within-subject variances ζ^2 and η^2 in the last two columns. The colorization for the two variances is different from the first three columns due to the range differences. For BGC (upper panel, two-tailed significance level p = 0.001), LME was slightly more powerful than SWB in small portion of the brain. For two-group comparison (n_1 = n_2 = 24, lower panel, two-tailed significance level p = 0.05), LME and SWT rendered very similar identifications. We note that 1) the parameters ρ, ζ^2, and η^2 for the group comparison (last three images in the lower panel) were estimated with the assumption of same variances across the groups (homoscedasticity); 2)multiple testing correction was not performed so that voxel-wise comparisons among the methods could be directly visualized; and 3) the three methods rendered virtually the same group estimate for ISC, but differed slightly in significance detection (i.e., the color at each

voxel is roughly the same across the three testing methods if the significance survives the corresponding threshold).

Table 1

Commands to process the anatomical data set using FreeSurfer, SUMA and AFNI prior to running *afni_proc.py.* The result of these steps is the production of tissue maps from each subject's anatomical volume to be used to create regressors for the FMRI processing.

```
# Commands run prior to afni_proc.py, each in appropriate
# directories for the data sets for each subject
# Run FreeSurfer on the anatomical, and then use
# SUMA to convert the FS output to NIFTI for AFNI to use.
recon-all -all -subject $subj -i $anat
@SUMA_Make_Spec_FS -sid $subj -NIFTI
# Select the ventricle maps from the FS output.
3dcalc -a aparc+aseg.nii -datum byte -prefix FT_vent.nii \
    -expr 'amongst(a,4,43)'
# Select the WM maps from the FS output.
3dcalc -a aparc+aseg.nii -datum byte -prefix FT_WM.nii \
    -expr 'amongst(a,2,7,16,41,46,251,252,253,254,255)'
```

Table 2

A compact tcsh script that contains the succint, tailored *afni_proc.py* command used to generate the full processing pipeline (> 500 lines) in AFNI for this study. To implement across the group, one simply loops through a list of subjects, entering the given file name as the sole command line argument, which is passed to the variable $subj.

```
# !/bin/tcsh
# FMRI processing script, ISC movie data.
# Assumes previously run FS and SUMA commands, respectively:
# $ recon-all -all -subject $subj -i $anat
# $ @SUMA_Make_Spec_FS -sid $subj -NIFTI
# Set top level directory structure
set subj = $1
set topdir = TOP_LEVEL_FILE_LOCATION
set task = movie
set fsroot = $topdir/freesurfer/subjects
set outroot = $topdir/subject_results/$task.6
# Input directory: unprocessed FMRI data
set indir = $ topdir/orig.data
# Input directory: FreeSurfer + @SUMA_MakeSpec_FS results
set fsindir = $fsroot/$subj/SUMA
# Output directory: name for output
set outdir = $outroot/$subj
# Input data: list of partitioned EPIs (movie clips)
set epi_dpattern = "movie*.HEAD"
# Input data: FreeSurfer results (anatomy, ventricle and WM maps)
set fsanat = ${subj}_SurfVol.nii
set fsvent = FSmask_vent.nii
set fswm = FSmask_WM.nii

afni_proc.py -subj_id $subj.$task                                  \
   -blocks despike tshift align tlrc volreg blur mask regress     \
   -copy_anat $fsindir/$fsanat                                     \
   -anat_follower_R0I aaseg anat $fsindir/aparc.a2009s+aseg_rank.nii  \
   -anat_follower_R0I aeseg epi $fsindir/aparc.a2009s+aseg_rank.nii   \
   -anat_follower_R0I FSvent epi $fsindir/$fsvent                  \
   -anat_follower_R0I FSWMe epi $fsindir/$fswm                     \
   -anat_follower_erode FSvent FSWMe                               \
   -dsets $epi_dpattern                                            \
```

```
-tcat_remove_first_trs 0                                \
-tlrc_base TT_N27+tlrc                                  \
-tlrc_NL_warp                                           \
-volreg_align_to MIN_OUTLIER                            \
-volreg_align_e2a                                       \
-volreg_tlrc_warp                                       \
-regress_ROI_PC FSvent 3                                \
-regress_make_corr_vols aeseg Fsvent                    \
-regress_anaticor_fast                                  \
-regress_anaticor_label FSWMe                           \
-regress_censor_motion 0.2                              \
-regress_censor_outliers 0.1                            \
-regress_apply_mot_types demean deriv                   \
-regress_est_blur_epits                                 \
-regress_est_blur_errts                                 \
-regress_run_clustsim no                                \
```