



## **Intraclass Correlations for Assessing Reliability in Presence of Within- and Between-Subject Variability in fMRI BOLD Responses**

Alexander Yu Zhigalov

Neuroscience Center, University of Helsinki – [a.zhigalov@gmail.com](mailto:a.zhigalov@gmail.com)

Summit Suen

Institute of Statistical Science, Academia Sinica – [summit.suen@gmail.com](mailto:summit.suen@gmail.com)

Juin-Der Lee

College of Commerce, National Chengchi University – [juinder.lee@gmail.com](mailto:juinder.lee@gmail.com)

Annabel S. H. Chen

College of HASS, Nanyang Technological University – [AnnabelChen@ntu.edu.sg](mailto:AnnabelChen@ntu.edu.sg)

Philip E. Cheng

Institute of Statistical Science, Academia Sinica – [pcheng@stat.sinica.edu.tw](mailto:pcheng@stat.sinica.edu.tw)

Michelle Liou\*

Institute of Statistical Science, Academia Sinica – [mliou@stat.sinica.edu.tw](mailto:mliou@stat.sinica.edu.tw)

The intraclass correlations (ICCs) assess the degree to which brain activation maps or connectivity networks resemble each other while subjects performing the same tasks twice or more. Theoretically, an ICC defines a ratio between two variances: one is the hypothetical true variance of outcomes across experimental replicates and the other is their observed variance. Although a variance ratio must lie in a range between zero and one, an observed ICC value can range from negative infinity to one because the true variance is estimated using empirical data. In this study, we considered a commonly used ICC index in fMRI studies, and derived its asymptotic standard error. Recently, the fluctuation in fMRI BOLD responses was found to be a better predictor to a subject's age and task performance compared to the average responses. In a sense, the within- and between-subject variances of ICC values might be spatially distributed in a pattern that was informative about developmental or aging effects in a group of participants having distinct demographic features. We also derived the ICC test statistics for assessing within- and between-subject reliability. Based on the test statistics, the supra-threshold voxels in an ICC map can be decided by controlling the false positive rate (FDR) using the voxel-FDR method based on the phase-randomization procedure or the cluster-FDR method based on standard results from the random field theory. In the empirical study, the ICC statistics were applied to assessing the within- and between-subject reliability of BOLD responses in an 8-min eyes-closed/open task (4-min each) with 30 healthy participants (15 males, average age  $22.50 \pm 3.462$ ). We made an empirical comparison between the voxel- and cluster-FDR methods. While the GLM and group independent component analysis suggested the occipital cortex to be more pronounced in the eyes-closed/open task, the supra-threshold voxels in the ICC maps indicated that BOLD responses in the thalamus and other subcortical regions were more reliable within individual subjects, and those in the precuneus were more reliable between subjects.

**Keywords:** asymptotic standard errors, BOLD signal, intraclass correlations, resting-state.