Concept Paper

Provisional Paper Title: Test-retest reliability of task-evoked BOLD fMRI in two samples: implications for individual differences neuroscience

Proposing Author: Annchen Knodt

Author's Email: ark19@duke.edu

P.I. Sponsor: Ahmad Hariri

Today's Date: 28 January 2019

Please describe your proposal in 2-3 pages with sufficient detail for helpful review.

Objective of the study:

Task-evoked activity in regions of interest (ROIs) measured with BOLD fMRI is increasingly used for mapping variability in behavior and risk for mental illness. However, the utility of neuroimaging phenotypes for the study of individual differences depends on how reliably they can be measured over time[1,2]. While studies assessing the test-retest reliability of MRI-derived measurements of many aspects of brain structure have found these to be highly reliable[3-5], evidence for lower test-retest reliability of task-based functional measures has begun to accumulate, indicating that these measures may not be suitable for individual differences research[6-8].

While most existing studies of test-retest reliability of regional fMRI task activation are limited to just one or two tasks, usually in a very small number of subjects, here we will leverage test-retest data from two datasets with a total of 11 tasks. In the Dunedin Longitudinal Study (DLS), 20 participants were scanned twice (mean test-retest interval of 79 days) while completing a battery of 4 fMRI paradigms targeting different cognitive domains: emotion / threat, motivation / reward, episodic memory, and executive control. In the Human Connectome Project (HCP), 45 participants were scanned twice (mean test-retest interval of approximately 140 days) while completing a battery of 7 fMRI paradigms: emotion / threat, reward / decision making, working memory / executive control, social cognition, higher order relational processing, language processing, and motor strip mapping[9].

A prominent metric used to quantify test-retest reliability in order to inform whether a measure is suitable for the study of individual differences is the intra-class correlation coefficient[10] (ICC), where a value of 1 indicates perfect reliability / stability over time, and values below 0.4 are considered "poor"[6]. Of the tasks to be examined here, some have been previously assessed for test-retest reliability of their target regions using the ICC, with either identical or highly similar variants of the task: the emotion / threat task (ICCs in the amygdala ranging from -0.02 to 0.7[7,11,12]), the reward tasks (ICCs in the ventral striatum ranging from 0.56 – 0.62[7,13]), the episodic memory task (ICCs in the hippocampus ranging from 0.59–0.87[14]), the executive control task (ICCs in the prefrontal cortex ranging from 0.42-0.52[15]), the working memory task (ICCs in the prefrontal cortex ranging from -0.4–0.77[7,16]), and several marginally similar variants of the motor task (ICCs in the motor cortex ranging from 0-0.85[6]). To our knowledge, test-retest ICCs have not been reported for the social cognition, relational processing, or language processing tasks.

Here we aim to 1) conduct the most comprehensive (to our knowledge) assessment to date of the test-retest reliability of regional activation using these 11 tasks and 2) discuss the implications of observed levels of test-retest reliability on the utility of regional measures of task activation in individual difference studies.

Data analysis methods:

To assess test-retest reliability of task-evoked activation and its implications in the use of these measures for the study of individual differences, we will conduct the following analyses for each of the 11 tasks:

1. Calculate group-level activation for the traditionally used condition of interest and ROI(s) for the respective task, to confirm engagement of the target region(s) by the task
2. Quantify the between-session reliability of the subject-level mean activation within the target ROI(s) using the intra-class correlation coefficient (ICC)
3. In order to further inform our interpretation of the reliabilities observed within the target ROI(s), quantify global reliability of activation using the whole-brain ICC, which has been suggested to be the strictest and most valuable measure of test-retest reliability[6]
4. To compare regional reliabilities for the given task, also calculate ICCs for the target ROIs from the other tasks, as well as for control regions in the visual cortex.
5. To provide further context and a comparison standard for evaluating the resulting task activation-related ICCs, calculate test-retest ICCs for structural measures including cortical thickness, cortical surface area, and sub-cortical volume.
6. Assess the relationship between reliability, sample size, and ability to detect effects of varying sizes at a given statistical power level.

Variables needed at which ages:

DLS Age 45 variables:

- Neuroimaging:
    - fMRI time course for the emotion / “Matching” task (targeting the amygdala)
    - fMRI time course for the executive control / “Colours” task (targeting the dorsolateral prefrontal cortex and the dorsal anterior cingulate cortex)
    - fMRI time course for the reward / “Quick-Strike” task (targeting the ventral striatum)
    - fMRI time course for the episodic memory / “Face-name” task (targeting the hippocampus)

HCP variables:

- fMRI time course for the emotion task (targeting the amygdala)
- fMRI time course for the reward task (targeting the ventral striatum)
- fMRI time course for the working memory task (targeting the dorsolateral prefrontal cortex)
- fMRI time course for the social / theory-of-mind task (targeting the lateral fusiform gyrus, superior temporal sulcus, and other “social-network” regions[17])
- fMRI time course for the relational task (targeting the rostrolateral prefrontal cortex[18])
- fMRI time course for the language task (targeting the anterior temporal lobe[19])
- fMRI time course for the motor task (targeting the motor cortex)

Significance of the Study (for theory, research methods or clinical practice):

This study will provide the most comprehensive assessment to date of test-rest reliability of fMRI regional activation, positioning it to serve as a widely applicable reference for understanding the place of task fMRI in the study of individual differences.

References cited:

1. Vul, Edward, Christine Harris, Piotr Winkielman, and Harold Pashler. 2009. “Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition.” Perspectives on Psychological Science 4:319–324.

2. Dichter, Gabriel S., Linmarie Sikich, Allen Song, James Voyvodic, and James W. Bodfish. 2012. “Functional neuroimaging of treatment effects in psychiatry: Methodological challenges and recommendations.” International Journal of Neuroscience 122:483–493.

3. Huang, Lejian et al. 2012. “Reproducibility of Structural, Resting-State BOLD and DTI Data between Identical Scanners.” PLoS ONE 7:1–8.

4. Wonderlick, J. S. et al. 2009. “Reliability of MRI-derived cortical and subcortical morphometric measures: Effects of pulse sequence, voxel geometry, and parallel imaging.” NeuroImage 44:1324–1333.

5. Vollmar, Christian et al. 2010. “Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners.” NeuroImage 51:1384–1394.

6. Bennett, Craig M., and Michael B. Miller. 2010. “How reliable are the results from functional magnetic resonance imaging?” Annals of the New York Academy of Sciences 1191:133–155.

7. Plichta, Michael M. et al. 2012. “Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery.” NeuroImage 60:1746–1758.

8. Infantolino, Zachary P., Katherine R. Luking, Colin L. Sauder, John J. Curtin, and Greg Hajcak. 2018. “Robust is not necessarily reliable: From within-subjects fMRI contrasts to between-subjects comparisons.” NeuroImage 173:146–152.

9. Barch, Deanna M. et al. 2013. “Function in the human connectome: Task-fMRI and individual differences in behavior.” NeuroImage 80:169–189.

10. Shrout, Patrick E, and Joseph L Fleiss. 1979. “Intraclass Correations: Uses in Assessing Rater Reliability.”

11. Johnstone T, Somerville LH, Alexander AL, Oakes TR, Davidson RJ, Kalin NH, et al. "Stability of amygdala BOLD response to fearful faces over multiple scan sessions." Neuroimage. 2005;25:1112-23.

12. Manuck, S.B., Brown, S.M., Forbes, E.E., Hariri, A.R., 2007. "Temporal stability of individual differences in amygdala reactivity." Am. J. Psychiatry 164, 1613–1614.

13. Fliessbach, K., Rohe, T., Linder, N.S., Trautner, P., Elger, C.E., Weber, B., 2010. "Retest reliability of reward-related BOLD signals." Neuroimage 50, 1168–1176.

14. Atri A, O'Brien JL, Sreenivasan A, et al. "Test-retest reliability of memory task functional magnetic resonance imaging in Alzheimer disease clinical trials." Arch Neurol. 2011;68(5):599-606.

15. Sheu LK, Jennings JR, Gianaros PJ. "Test-retest reliability of an fMRI paradigm for studies of cardiovascular reactivity. Psychophysiology." 2012;49:873-84.

16. Caceres, Alejandro, Deanna L. Hall, Fernando O. Zelaya, Steven C.R. Williams, and Mitul A. Mehta. 2009. “Measuring fMRI reliability with the intra-class correlation coefficient.” NeuroImage 45:758–768.

17. Wheatley, Thalia, Shawn C Milleville, and Alex Martin. 2007. “Understanding Animate Agents: Distinct Roles for the Social Network and Mirror System.” Psychological Science 18:469–474.

18. Smith, Rachelle, Kamyar Keramatian, and Kalina Christoff. 2007. “Localizing the rostrolateral prefrontal cortex at the individual level.” NeuroImage 36:1387–1396.

19. Binder JR, Gross WL, Allendorfer JB, Bonilha L, Chapin J, Edwards JC, Grabowski TJ, Langfitt JT, Loring DW, Lowe MJ, Koenig K, Morgan PS, Ojemann JG, Rorden C, Szaflarski JP, Tivarus ME, Weaver KE. "Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study." Neuroimage. 2011;54:1465–1475.

Data Security Agreement

Provisional Paper Title	Test-retest reliability of task-evoked BOLD fMRI in two samples: implications for individual differences neuroscience
Proposing Author	Annchen Knodt
Today's Date	28 January 2019

Please keep one copy for your records and return one to the PI Sponsor

Please initial your agreement

ARK	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
ARK	My project is covered by Duke or Otago ethics committee OR I have /will obtain ethical approval from my home institution.
ARK	I will treat all data as "restricted" and store in a secure fashion. My computer or laptop is: a) encrypted (recommended programs are FileVault2 for Macs, and Bitlocker for Windows machines) b) password-protected c) configured to lock-out after 15 minutes of inactivity AND d) has an antivirus client installed as well as being patched regularly.
ARK	I will not "sync" the data to a mobile device.
ARK	In the event that my laptop with data on it is lost, stolen or hacked, I will immediately contact Professor Moffitt or Caspi. (919-684-6758, tem11@duke.edu, ac115@duke.edu)
ARK	I will not share the data with anyone, including my students or other collaborators not specifically listed on this concept paper.
ARK	I will not post data online or submit the data file to a journal for them to post. *Some journals are now requesting the data file as part of the manuscript submission process. The Dunedin Study Members have not given informed consent for unrestricted open access, so we have a managed-access process. Speak to Terrie or Avshalom for strategies for achieving compliance with data-sharing policies of journals.*
ARK	I will delete all data files from my computer after the project is complete. Collaborators and trainees may not take a data file away from the office. The data remains the property of the Study and cannot be used for further analyses without an approved concept paper for new analyses.

Signature: