Appendix Section B(2): CONCEPT PAPER TEMPLATE

DUNEDIN MULTIDISCIPLINARY HEALTH AND DEVELOPMENT STUDY

(The Dunedin Study)

CONCEPT PAPER TEMPLATE

(July 2024)

DUNEDIN STUDY CONCEPT PAPER

Provisional Paper Title: Test-Retest Reliability of Midlife Change in Brain Structure

Proposing Author: Annchen Knodt

Author's Email: ark19@duke.edu

P.I. Sponsor: Ahmad Hariri
(if the proposing author is a student or colleague of an original PI)

Today's Date: 11/8/24

Please describe your proposal in 2-3 pages with sufficient detail for helpful review by addressing all areas outlined below.

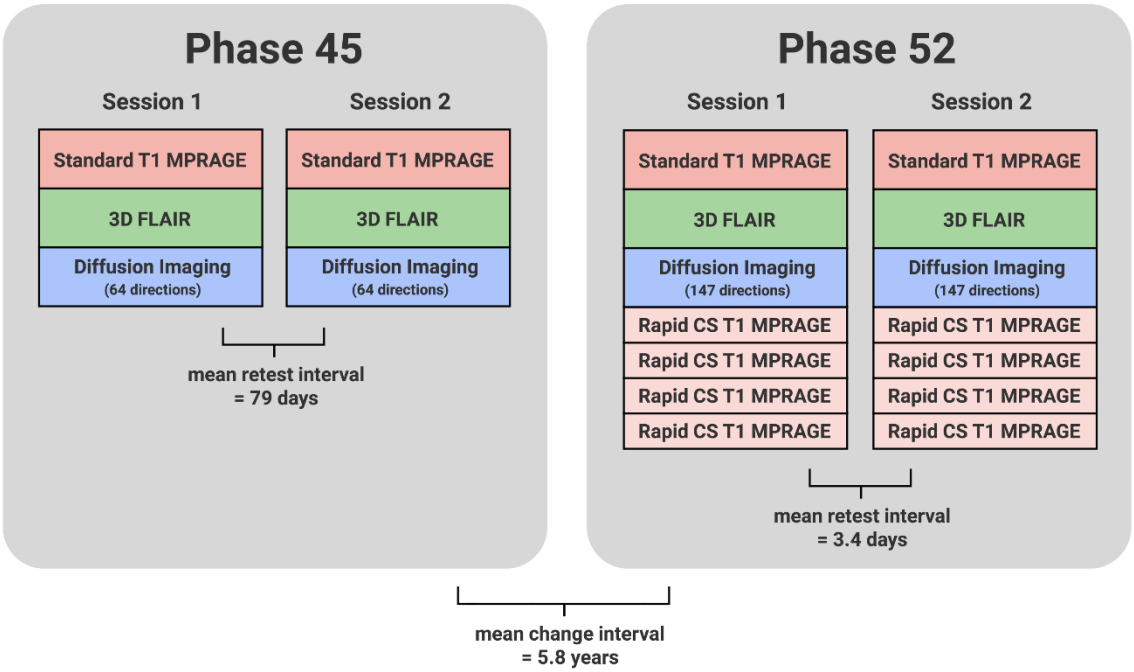
Objective of the study:

Longitudinal studies of within-individual change in MRI-based measures of brain structure are important for understanding how brain changes shape cognitive functioning and brain-based disease risk over time[1-3]. However, our ability to detect replicable associations between changes in brain structure and such outcomes depends on the reliability of our MRI-based measures of change[4]. Given that measures of change broadly may be inherently susceptible to reductions in reliability[5,6], it is especially important to understand the psychometric properties of our MRI-based measures of brain structural changes and their implications for our ability to detect replicable associations in studies of individual differences.

The gold standard way to assess reliability empirically is by running identical experiments on two separate occasions, and computing the correspondence, or "test-retest" reliability, between the two sessions using the intraclass correlation coefficient (ICC). As such, the empirical evaluation of change reliability requires two identical experiments at both ends of the interval across which the change is measured. Due to the substantial investment required to conduct these additional experiments, especially in areas of research such as neuroimaging, these studies are rarely undertaken.

Indeed, only a small number of studies to date have sought to evaluate the reliability of MRI-based measures of changes in brain structure. One study examined the test-retest reliability of two-year change in voxel-based morphometry (VBM) measures using within-session repeated 1.5T scans in older adults, finding good to excellent reliability, albeit slightly lower in patients with mild cognitive impairment and Alzheimer's disease compared to healthy volunteers[7]. In a similar but independent sample, the same research group also reported good replicability of two-year change in VBM as well as cortical thickness that also differed between diagnostic groups[8,9]. Lastly, a recent study evaluated the reliability of MRI-based structural change measures by building models with parameters derived from cross-sectional and single time point test-retest data, finding higher reliability for 1) subcortical verses cortical measures, 2) older versus younger adults, and 3) longitudinal versus cross-sectional data processing protocols[10]. To build a more complete understanding of the reliability of MRI-based measures of changes in brain structure, further multi-session empirical test-retest studies across the lifespan are needed.

We will try to help fill this gap by examining the test-retest reliability of 6-year midlife changes in commonly used MRI-based measures of brain structure in 20 members of the Dunedin Study who underwent two MRI sessions at both the 45- and 52-year assessment Phases with between-session intervals averaging 79 days and 3 days, respectively (see Figure below for a schematic of the study design). We will use the four MRI sessions to assess test-retest reliability of change from Phase 45 to Phase 52 across cortical and subcortical structural measures derived from T1-weighted scans as well as measures of white matter integrity derived from diffusion-weighted scans. We will assess change reliability for these measures using different data processing protocols and conduct sensitivity analyses using different schemes for parcellation of the cortex.



We note that the reliability of MRI-based measures of changes in brain structure is likely to depend on many factors, including 1) the reliability of the measures at each timepoint, 2) the length of the retest interval at each time point, 3) the length of interval between the two time points, 4) the age(s) and developmental stage(s) at which measurements are taken, 5) the specific brain region and feature being measured, and 6) the amount of between-individual variability in the test-retest sample. Our study features 1) measurements with excellent single time point test-retest reliability (most ICCs > .9), 2) a retest interval averaging 79 days at Phase 45 and 3 days at Phase 52, 3) an average interval between phases of 5.8 years, 4) measurements taken between the ages of 45 and 52 where developmental and aging-related changes are minimal, 5) a variety of gray and white matter measures across the brain, and 6) a sample drawn from a population-representative both cohort that matches the original sample on most demographic indicators. Thus, we are able to shed light on the reliability of MRI-based measures of changes in brain structure under this set of conditions, and further studies will be needed to fully explore the effects of each factor.

Data analysis methods[1]:

First, structural MRI data will be processed using four different FreeSurfer processing streams: 1) cross-sectional processing employing a standard T1 mprage at Phase 45 and Phase 52, with a 3D FLAIR to improve grey matter boundary detection, 2) longitudinal processing employing the same T1 mprage and 3D FLAIR scans, 3) longitudinal processing employing only the T1 mprage scans, and 4) longitudinal processing with the T1 and FLAIR scans and an additional four "rapid" (compressed sensing) T1 scans collected at Phase 52 to improve measurement precision. For the longitudinal processing streams, data from test and retest sessions will be processed completely independently (i.e., separate template steps). Additionally, diffusion-weighted imaging data from both sessions at both phases will be processed using an FSL-based diffusion tensor imaging protocol to compute fractional anisotropy (FA)[11], an indicator of white matter microstructural integrity.

Second, specific MRI-based structural measures of interest will be extracted from the pre-processed data from both sessions at both phases. To explore the reliability of changes in global measures of broad interest and of particular relevance for aging, we will begin by focusing on total gray matter volume, mean cortical thickness, and total surface area, as well as the mean volumes of the hippocampus, lateral ventricles, and white matter hypointensities estimated by FreeSurfer. To explore regional differences in change reliability, we will additionally extract regional measures of cortical gray matter thickness, surface area, and volume using FreeSurfer's Desikan-Killiany atlas[12] as well as subcortical gray matter volumes using the "aseg" atlas[13]. Finally, we will extract global and regional measures of fractional anisotropy using the JHU atlas[14].

Next, for each of the extracted brain measures, and for each of the two test-retest sessions, we will compute annualized percentage change from Phase 45 to Phase 52. We will then compute the intraclass correlation coefficient (ICC) between change scores from the two sessions to assess the reliability of change.

Finally, since previous studies have reported differences in longitudinal stability across alternative cortical parcellations[15], we will conduct sensitivity analyses using the more fine-grained Destrieux parcellation[16] and explore the relationship between change reliability and cortical region size using Pearson correlations.

Variables needed at which ages:

Phase 45: T1, FLAIR, and diffusion weighted scans from 20 test-retest Study members; sex; exact age at scan

Phase 52: T1, FLAIR, diffusion weighted, and 4 compressed sensing T1 scans from 20 test-retest Study members; sex; exact age at scan

[1] A key concern for the Dunedin Study is superficial analyses of data that simply identify differences or deficits between ethnic groups or other communities where inequities exist (e.g. persons with disabilities, Pasifika peoples, members of migrant and SOGIESC (Sexual Orientation, Gender Identify and Expression and Sexual Characteristics) communities). The cumulative effect of these types of studies is stigmatising and not of benefit. Any research that identifies differences must (a) incorporate information on the broader context (e.g. historical or political factors); (b) where possible undertake additional analyses to examine the source of the difference/s, and (c) include policy recommendations for its resolution.

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

Establishing the reliability of MRI-based measures of change in brain structure is critical for their application in studying individual differences in trajectories of cognitive functioning and brain-based disease risk. The proposed research will help to fill a substantial gap in the evaluation of the reliability of such measures.

How the paper will contribute to Māori health advancement and/or equitable health outcomes[2]

While this study does not include any ethnicity-focused analyses, a better understanding of the reliability of MRI-based measures of changes in brain structure will lead to broad improvements in the replicability of research in this area generally. This will ultimately result in improved diagnosis, treatment, and prevention of brain-related illnesses, to which socially disadvantaged individuals are particularly susceptible.[17,18]

References:

1. Akshoomoff N, Newman E, Thompson WK, et al. The NIH Toolbox Cognition Battery: results from a large normative developmental sample (PING). *Neuropsychology*. Jan 2014;28(1):1-10. doi:10.1037/neu0000001
2. Lindenberger U, von Oertzen T, Ghisletta P, Hertzog C. Cross-sectional age variance extraction: what's change got to do with it? *Psychol Aging*. Mar 2011;26(1):34-47. doi:10.1037/a0020525
3. Maxwell SE, Cole DA. Bias in cross-sectional analyses of longitudinal mediation. *Psychol Methods*. Mar 2007;12(1):23-44. doi:10.1037/1082-989X.12.1.23
4. Nunnally JC. *Introduction to psychological measurement*. McGraw-Hill; 1959.
5. Cronbach LJ, Furby L. How we should measure" change": Or should we? *Psychological bulletin*. 1970;74(1):68.
6. Zimmerman DW. The reliability of difference scores in populations and samples. *Journal of Educational Measurement*. 2009;46(1):19-42.
7. Takao H, Amemiya S, Abe O, Alzheimer's Disease Neuroimaging I. Reliability of Changes in Brain Volume Determined by Longitudinal Voxel-Based Morphometry. *J Magn Reson Imaging*. Aug 2021;54(2):609-616. doi:10.1002/jmri.27568
8. Takao H, Amemiya S, Abe O, Alzheimer's Disease Neuroimaging I. Reproducibility of Brain Volume Changes in Longitudinal Voxel-Based Morphometry Between Non-Accelerated and Accelerated Magnetic Resonance Imaging. *J Alzheimers Dis*. 2021;83(1):281-290. doi:10.3233/JAD-210596
9. Takao H, Amemiya S, Abe O, Alzheimer's Disease Neuroimaging I. Reproducibility of Longitudinal Changes in Cortical Thickness Determined by Surface-Based Morphometry Between Non-Accelerated and Accelerated MR Imaging. *J Magn Reson Imaging*. Apr 2022;55(4):1151-1160. doi:10.1002/jmri.27929
10. Vidal-Pineiro D, Sorensen O, Stromstad M, et al. Reliability of structural brain change in cognitively healthy adult samples. *bioRxiv*. 2024:2024-06.
11. Jahanshad N, Kochunov PV, Sprooten E, et al. Multi-site genetic analysis of diffusion

[2] Helpful information can be found here: https://www.hrc.govt.nz/sites/default/files/2020-01/NZ%20Prioritisation-Framework-FA-web_0.pdf

images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. *Neuroimage*. Nov 1 2013;81:455-469. doi:10.1016/j.neuroimage.2013.04.061
12. Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. Jul 1 2006;31(3):968-80. doi:10.1016/j.neuroimage.2006.01.021
13. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. Jan 31 2002;33(3):341-55. doi:10.1016/s0896-6273(02)00569-x
14. Mori S, Wakana S, van Zijl PCM, Nagae-Poetscher LM. *MRI atlas of human white matter*. Elsevier Science; 2005.
15. Parsons S, Brandmaier AM, Lindenberger U, Kievit R. Longitudinal stability of cortical grey matter measures varies across brain regions, imaging metrics, and testing sites in the ABCD study. *Imaging Neuroscience*. 2024;2:1-22.
16. Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*. Oct 15 2010;53(1):1-15. doi:10.1016/j.neuroimage.2010.06.010
17. Chen-Xu J, Varga O, Mahrouseh N, et al. Subnational inequalities in years of life lost and associations with socioeconomic factors in pre-pandemic Europe, 2009–19: an ecological study. *The Lancet Public Health*. 2024;9(3):e166-e177.
18. Marmot MG, Smith GD, Stansfeld S, et al. Health inequalities among British civil servants: the Whitehall II study. *Lancet*. Jun 8 1991;337(8754):1387-93. doi:10.1016/0140-6736(91)93068-k