Test-Retest Consistency of Hippocampal Subfield Volume Measures: Implications for Longitudinal Developmental Studies

Roya Homayouni, Qijing Yu, Sruthi Ramesh, Ana Daugherty, Noa Ofen

Institute of Gerontology & Department of Psychology, Wayne State University

Ofen Lab
Cognitive and Brain Development

Introduction

The hippocampus (Hc), a fundamental neural substrate of memory, is composed of cytoarchitectonically distinct subfields: dentate gyrus (DG), cornu ammonis sectors (CA1-3) and subiculum [1], thought to support distinct memory aspects [4].

Evidence of differential maturation across the Hc subfields [2] suggests that Hc subfields maturation underlies memory development [3,4]. This possibility can be tested with longitudinal structural MRI studies, with reliable measurement of Hc subfield volumes, such as that achieved by manual demarcation.

To interpret differential development of these regions, error variance must be the same across subfields. Yet, Hc subfields structure can introduce potentially differential source of errors due to their distinct morphological and volumetric properties. Importantly, the errors may show age related bias in developmental samples due to the factors such as differences in movement or head size across visits. Manual demarcation additionally may exhibit a degree of variability over time and between raters, leading to inconsistency of Hc subfields measurement between time points[5].

Here, we assessed the test-retest consistency of Hc subfields volume measures obtained with manual demarcation of ultra-high Hc MR images acquired at different time points. **We established a high longitudinal consistency of Hc subfield volumes using our methods, a prerequisite for interpreting meaningful longitudinal changes.**

Methods

Participants

Sample 1: two MRIs, one-month delay (n=28, ages 7-20 years, mean=12.64, SD=3.35)

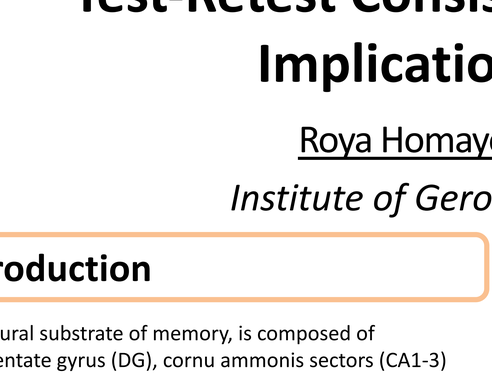
Sample 2: two MRIs, two-year delay (n=28, ages 8-17 years, mean=11.90, SD=2.78)

Brain Imaging

High-resolution structural MR images (T2 PD TSE, 0.4x0.4x2.0 mm^3) were collected with a 3T Scanner at WSU

Hippocampal Subfields Volume

Manual demarcation of the subfields were conducted by raters with high inter-rater reliability (ICC(2) > 0.85, ICC(3) > 0.90)



Longitudinal Protocol

A randomized procedure was implemented to blind raters to the information about ID and timepoints. To make consistent decision across two timepoints, the brain pairs were traced simultaneously and slice by slice

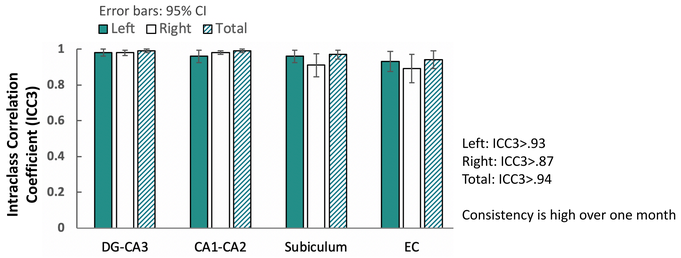
Statistical Analyses

Test-retest Reliability Using Intraclass Correlation Coefficient: Two way mixed-effect single ANOVA model that assumes non-independent observation (ICC3,1; Shrout and Fleiss, 1979 [7]) was conducted for Hc subfield volumes across two visits

Bias Evaluation*:* To determine if measurement error correlates with size of Hc subfields we implemented Bland-Altman Plots. Pearson correlation and one sample t-test were conducted to assess the measurement bias. Bar-graphs and boxplots were plotted to visualize the ICC(3) measures and sensitivity of the method in capturing individual differences, respectively. Error bars represent 95% confidence intervals.

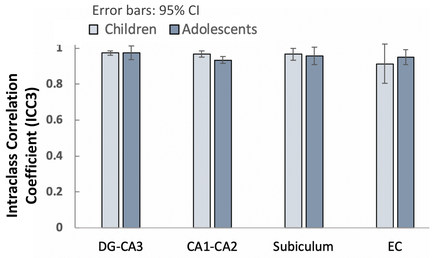
Results

High Consistency of Hc Subfield Volumes: One-Month Delay



Left: ICC3>.93
Right: ICC3>.87
Total: ICC3>.94

Consistency is high over one month

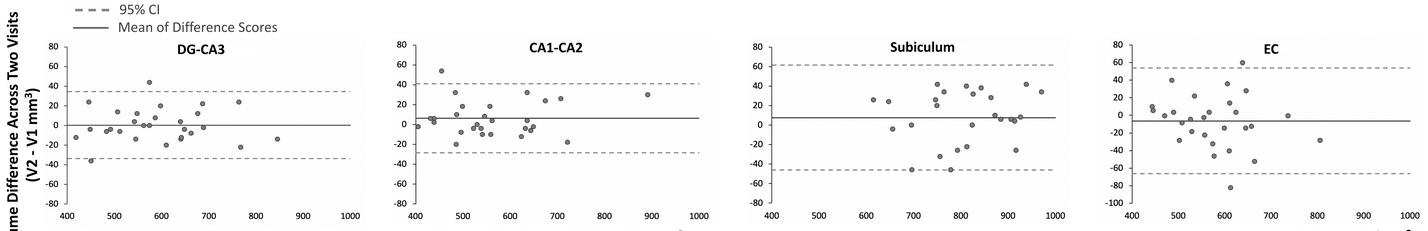


Children (age 7-12 years)
Left: ICC3>.91
Right: ICC3>.84
Total: ICC3>.91

Adolescents (age 13-20 years)
Left: ICC3>.93
Right: ICC3>.90
Total: ICC3>.96

Children and adolescents did not differ in consistency over one month

No Systematic Bias on Hc Subfield Volumes Estimation Over One-Month Delay



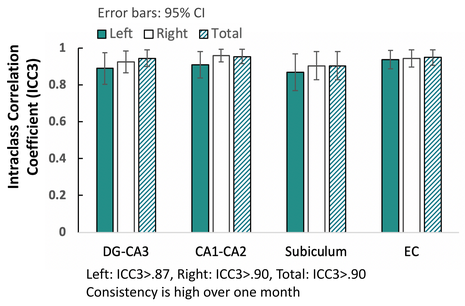
Average Volumes of V1 and V2 (mm^3)
Left: r = .12, p = .55; t= -.52, p = .60
Right: r = -.06, p = .77; t= .71, p = .48
Total: r = .02, p = .90; t= .11, p = .91

Average Volumes of V1 and V2 (mm^3)
Left: r = -.02, p = .92; t= 1.8, p = .07
Right: r = .12, p = .56; t= .51, p = .61
Total: r = .10, p = .63; t= 1.89, p = .07

Average Volumes of V1 and V2 (mm^3)
Left: r = -.26, p = .18; t= 1.3, p = .18
Right: r = .16, p = .42; t= .68, p = .50
Total: r = .06, p = .77; t= 1.5, p = .14

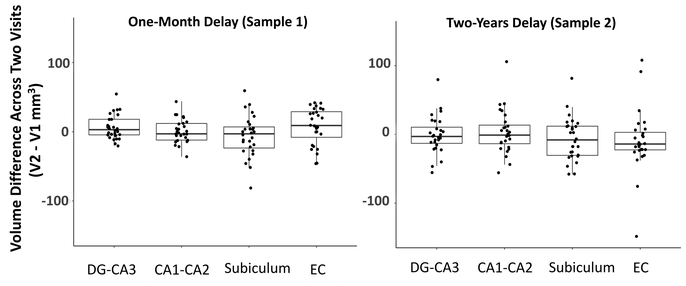
Average Volumes of V1 and V2 (mm^3)
Left: r = -.42, p = .03*; t= -.28, p = .78
Right: r = .03, p = .88; t= -1.27, p = .22
Total: r = -.18, p = .35; t= -1.09, p = .28
*Not significant after FDR correction

High Consistency: Two-Year Delay



Left: ICC3>.87, Right: ICC3>.90, Total: ICC3>.90
Consistency is high over one month

Larger Individual Variability over Two-Year Delay



Variance of Difference Scores for Total Volume [95% CI]

One-Month Delay
DG-CA3: 288.60 [155.58-417.26]
CA1-2: 302.91 [161.55-447.77]
Subiculum:719.91 [465.97-920.06]
EC:897.286 [472.9-1295.94]

Two-Year Delay
DG-CA3: 1010.81[454.26-1629.94]
CA1-2: 740.57 [360.11-1102.66]
Subiculum: 2027.07 [818.93-3257.13]
EC: 983.24 [558.91-1414.53]

Summary

Over one month, assessing test-retest reliability of Hc subfield measures obtained with manual tracing, we found:

- excellent consistency between the two visits
- consistency did not differ across age
- no systematic bias based on volume measures between the two visits

Over two years, we found excellent consistency between Hc subfields volume measures, yet, compared to over one month, there was larger individual variability in DG-CA3 volumes, indicating high sensitivity of our method

These findings support excellent consistency between Hc subfield volume measures assessed by our manual demarcation protocol and MRI acquisition parameters.

References:

1. *Jones, M. W., McHugh, T. J., (2011). Updating hippocampal representations: CA2 joins the circuit. Trends Neurosci 34:526–535. doi: 10.1016/j.tins.2011.07.007*
2. *Lavenex, P., Lavenex P. B., (2013). Building hippocampal circuits to learn and remember: Insights into the development of human memory. Behav Brain Res 254:8–21*
3. *Daugherty, A. M., Flinn, R. W., Ofen, N. (2017). Age-related differences in CA3-dentate gyrus volume uniquely linked to improvement in associative memory from childhood to adulthood. NeuroImage. 153: 75-85.*
4. *Keresztes, A., Ngo, C. T., Lindenberger, U., Werkle Bergner, M., Newcombe. N.S. (2018). Hippocampal maturation drives memory from generalization to specificity. Trends Cogn. Sci., 22 (2018), pp. 676-686*
5. *Yushkevich P.A. Pluta J.B. Wang H. Xie L. Ding S.L. Gertje E.C. Mancuso L. Kliot D. Das S.R. Wolk D.A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. Hum. Brain Mapp; 36: 258-287*
6. *Shrout PE, Fleiss JL (1979): Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 86:420–428*