Translational MRI Volumetry with NeuroQuant: Effects of Version and Normative Data on Relationships with Memory Performance in Healthy Older Adults and Patients with Mild Cognitive Impairment

Julija Stelmokas, *Veterans Affairs Ann Arbor Healthcare System*
Lance Yassay, *Veterans Affairs Ann Arbor Healthcare System*
Bruno Giordani, *University of Michigan*
Hiroko H. Dodge, *University of Michigan*
Ivo D. Dinov, *University of Michigan*
Arijit Bhaumik, *University of Michigan*
Krish Sathian, *Emory University*
Benjamin Hampstead, *Emory University*

Translational MRI volumetry with NeuroQuant: Effects of version and normative data on relationships with memory performance in healthy older adults and patients with mild cognitive impairment

Julija Stelmokas[1], **Lance Yassay**[1,2], **Bruno Giordani**[2,3,4], **Hiroko H. Dodge**[4,5], **Ivo D. Dinov**[6,7,8], **Arijit Bhaumik**[4], **K. Sathian**[9,10], and **Benjamin M. Hampstead**[1,2]

[1]Veterans Affairs Ann Arbor Healthcare System, Mental Health Service (116B), 2215 Fuller Road, Ann Arbor, MI 48105, United States of America

[2]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

[3]Departments of Neurology, and Psychology and School of Nursing, University of Michigan, Ann Arbor, MI, USA

[4]Michigan Alzheimer's Disease Core Center, University of Michigan, Ann Arbor, Michigan, United States of America

[5]Department of Neurology and Layton Aging and Alzheimer's Disease Center, Oregon Health &Science University, USA

[6]Statistics Online Computational Resource, School of Nursing, Michigan Institute for Data Science, University of Michigan, Ann Arbor, Michigan, United States of America

[7]Stevens Neuroimaging and Informatics Institute, University of Southern California, Los Angeles, California, United States of America

[8]Udall Center of Excellence for Parkinson's Disease Research, University of Michigan, Ann Arbor, Michigan, United States of America

[9]Departments of Neurology, Rehabilitation Medicine, and Psychology, Emory University, Atlanta, GA, USA

[10]Rehabilitation R&D Center for Visual and Neurocognitive Rehabilitation, Atlanta VAMC, Decatur, GA, USA

Abstract

NeuroQuant (NQ) is a fully-automated program that overcomes several existing limitations in the clinical translation of MRI-derived volumetry. The current study characterized differences between the original (NQ1) and an updated NQ version (NQ2) by (i) replicating previously identified relationships between neuropsychological test performance and medial temporal lobe volumes, (ii) evaluating the level of agreement between NQ versions, and (iii) determining if the addition of

Corresponding author: Benjamin M. Hampstead, Ph.D. 2101 Commonwealth Blvd Suite C, Ann Arbor, MI 48105, Phone: 734-763-9259, Fax: 734-936-8262, bhampste@med.umich.edu.

NQ2 age-/sex-based z-scores hold greater clinical utility for prediction of memory impairment than standard percent of intracranial volume (%ICV) values. Sixty-seven healthy older adults and 65 MCI patients underwent structural MRI and completed cognitive testing, including the Immediate and Delayed Memory indices from the RBANS. Results generally replicated previous relationships between key medial temporal lobe regions and memory test performance, though comparison of NQ regions revealed statistically different values that were biased toward one version or the other depending on the region. NQ2 hippocampal z-scores explained additional variance in memory performance relative to %ICV values. Findings indicate that NQ1/2 medial temporal lobe volumes, especially age- and sex-based z-scores, hold clinical value, though caution is warranted when directly comparing volumes across NQ versions.

Keywords

hippocampus; amygdala; Alzheimer's disease; memory; neuroimaging

INTRODUCTION

MRI-derived volumetry is a well-established biomarker of Alzheimer's disease (AD) [1–3] and can potentially aid in clinical staging [4]. However, attempts to translate volumetrics into clinical practice are limited by a number of pragmatic challenges, including inadequate access due to the need for specialized expertise and sufficiently powered equipment, lengthy processing time, non-standardized (laboratory specific) approaches, and the need for a normative context within which to consider an individual patient's values (see discussions in [5] and [6]). NeuroQuant (NQ) may bridge this translational gap by overcoming several of the aforementioned limitations. Specifically, NQ is a commercially available (i.e., widespread access) and fully automated service that uses a standardized processing pipeline to provide volumetrics in a clinically-friendly time period of about 15 minutes. NQ may prove affordable for clinical implementation, since its FDA approval (510(k) K061855) allows providers to bill for its use. Perhaps most importantly, NQ contextualizes volumes for each region by providing age- and sex-based normative values (in z-scores) that are based on large-scale MRI databases [7].

Results from several studies indicate that NQ provides a reliable and valid measurement of brain volumes in clinical and healthy samples compared to computer aided manual segmentation [8], visual inspection [9] and other fully automated programs [10]. NQ volumes also appear to have clinical relevance. For instance, Kovacevic and colleagues [11] found that NQ-derived medial temporal lobe volumes correlated with memory performance in patients with mild cognitive impairment (MCI). However, one limitation of this study was that brain volumes were adjusted for percentage of whole brain volume rather than intracranial volume, which may have reduced reliability [12]. We previously reported that intracranial volume-corrected inferior lateral ventricle volumes were significantly correlated with immediate and delayed memory performance, and that hippocampal volumes were correlated with delayed memory performance on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) in a mixed sample of MCI and healthy older adults (HOA) [13]. Importantly, the relationship between hippocampus and delayed memory

persisted in the MCI subgroup, further reinforcing the potential clinical utility of the data. While these findings are encouraging, NQ underwent a substantial revision in mid-2015 (version 2 – NQ2), including an updated segmentation approach and expanded age range [14]. Differences may exist between the original NQ1 and NQ2 values as well as their relationship with memory test performances that have yet to be examined, especially using independent datasets.

The current study used MRI data from an expanded sample, relative to our earlier report by England and colleagues [13], which consisted of cognitively intact older adults and those with MCI. Our first aim was to replicate previous findings demonstrating the relationship between key medial temporal lobe regions and cognitive test performance. Second, we aimed to directly compare the two NQ versions (i.e., NQ1 and NQ2) in order to identify and characterize any differences that could affect the interpretation and clinical translation of older research findings. Finally, we sought to extend findings by comparing the relationship between memory test performance and available medial temporal lobe volumes using the standard research-based normalization method of intracranial volume (ICV) correction and the clinically-friendly age- and sex-based z-scores.

MATERIALS AND METHODS

Participants

Neuropsychological and MRI data from a total of 132 right-handed participants were used in the current study, 100 of whom were included in our previous report [13]. Of the total sample, 67 were classified as HOA and 65 were classified as MCI. We excluded data from an additional 3 participants whose neuropsychological profiles were inconsistent with their clinical diagnoses. All participants were recruited as part of a larger research program that evaluated structural and functional neuroimaging changes following cognitive rehabilitation interventions in healthy aging and MCI. MCI patients were recruited from the Emory Alzheimer's Disease Research Center and affiliated memory disorder clinics (both at Emory University and the Atlanta VA Medical Center). Diagnosis was based on Petersen's 2004 criteria [15] with 1) subjective memory complaints (provided by the patient or an informant), 2) objective evidence of memory decline, and 3) preserved everyday functioning and arose during a consensus conference of neurologists, geriatricians, and/or neuropsychologists. Healthy older adults were recruited from the community and these same resources (e.g., Emory ADRC registries, patient spouses). All individuals underwent additional neuropsychological testing at the time of enrollment into our studies; these data were independent from those used in making the initial diagnosis. All data were collected under studies approved by the Institutional Review Board of Emory University and the Research and Development Committee of the Atlanta VAMC. Each participant provided written informed consent. Basic demographic and neuropsychological data from the time of study enrollment can be seen in Table 1.

Neuropsychological assessment

All subjects completed a brief neuropsychological assessment that included the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; [16]), Trail-Making Test

(TMT), Part A and B [17], and the Mini-Mental State Examination (MMSE) [20]. The 15-item Geriatric Depression Scale (GDS-15) was administered to rule out depression [18] and the Functional Assessment Questionnaire (FAQ) [19] was administered to ensure that functional status had remained largely intact. These data were acquired at the time of study enrollment, which was generally within one month of MRI scanning (see below), and were used to ensure that MCI patients had not progressed or reverted, and that HOA remained cognitively intact.

Magnetic Resonance Imaging (MRI) acquisition and analysis

Magnetic resonance imaging was performed on all participants with a 3T Siemens Trio MRI Scanner (Siemens Medical Solutions, Malvern, PA) using a 12-channel head coil. High-resolution anatomic scans were collected using a 3D MPRAGE sequence (TR 2300 ms, TE 3.9 ms, FA 8°, FOV 256mm, IPR 1×1mm, IPM 256×256) with 176 sagittal slices (1 mm thick). An MR technologist reviewed scans for quality at the time of acquisition and repeated/replaced those of poor quality as necessary.

Volumetric analyses

NQ analyses were completed for all participants using both NQ1 and NQ2 via the standard processing pipeline (i.e., uploading raw DICOM files to the secure NQ servers for processing). The details of this procedure are previously described elsewhere [8, 13]. Briefly, the protocol includes a quality check, correction for gradient non-linearity/B1 field inhomogeneity, and skull stripping followed by a discrete cosine transformation and registration onto a probabilistic atlas. An anatomic label is assigned to each voxel based on estimates from the probabilistic atlas.

Both NQ1 and NQ2 provide a report that includes raw and corrected volumes (based on %ICV) for 11 brain regions. As in our earlier study [13], we selected the hippocampus and inferior lateral ventricle for correlations with cognitive test performance. We also included the amygdala volume as an additional region of interest, given previous research implicating its relationship to memory performance and MCI [21]. In order to extend research findings using NQ2, we correlated volumetric z-scores for the hippocampus and amygdala (inferior lateral ventricle normative data is not provided by NQ2) with memory test performance. NQ2 also provides normative (z-score) values for the isthmus and "post-cingulate" region (note this region comprises the posterior cingulate cortex according to a personal communication with CorTechs Labs [22]), which we included given evidence of reduction in those with MCI [23] and utility in predicting conversion from MCI to probable AD [24].

Statistical analyses

IBM SPSS version 24 was used to perform all statistical analyses. Data met parametric assumptions based on visual-inspection and observed skewness coefficients <2.0 [25]. As appropriate, the differences in demographics between groups (MCI versus HOA) were evaluated using chi-square test or independent samples t-test. The results indicated a significant difference in sex between the groups. Therefore, a multivariate analysis of covariance (MANCOVA) (controlling for sex, with post-hoc tests as appropriate) was used to assess the differences between groups in ICV (NQ1 and NQ2) and neuropsychological

test performance. Independent sample *t*-tests were used to evaluate differences between z-score NQ2 volumes, since sex was already controlled for in these values.

In order to address our first aim, we utilized partial correlations (controlling for effect of sex) to evaluate the relationship between selected structures (total hippocampus, inferior lateral ventricle, and amygdala) and cognitive test performance across the entire sample and subgroups (MCI and HOA). Although we were primarily interested in memory test performances, we included other RBANS indices to evaluate the specificity of the presumed MTL relationships (i.e., discriminant validity). The false discovery rate (FDR) [28] was used to correct for multiple comparisons (computed P(#tests+1))/(2*(#tests)).

To address our second aim of comparing NQ versions, we computed Bland Altman (B&A) plots [26–27], which are a graphical method of comparing the differences between two methods (NQ1 vs. NQ2) against the averages of the two paired measurements (NQ1 vs. NQ2). Results of the B&A plots provide estimates of level of agreement between NQ1 and NQ2. A one-sample t-test of the difference between the two versions (NQ1 - NQ2) was calculated for each volume to determine if the difference was significantly different from zero and to evaluate for bias (i.e., if either NQ version provided larger or smaller values than the other). There was no *a priori* clinically acceptable difference identified because our primary aim was to evaluate whether NQ versions were statistically biased, in general. We used a 95% confidence interval to establish the level of agreement. For this aim, we did not control for multiple comparisons given the greater risk of falsely concluding that the measures were comparable (i.e., *p*-values that fail to surpass the FDR correction threshold), which could result in the inappropriate use of cutoffs/values for diagnostic or treatment planning purposes.

To address our third aim of evaluating the clinical utility of NQ2 (z-scores) versus standard NQ2 (%ICV) values, we utilized partial correlations that controlled for the effects of shared variance between the NQ2 volumes. First, we computed a partial correlation between NQ2 z-score values (hippocampus, amygdala) and memory test performance and controlled for the effects of NQ2 (%ICV). Then, we computed a separate partial correlation between NQ2 (%ICV) (hippocampus, amygdala) and memory test performance and controlled for the effects of NQ2 z-score values. FDR correction was used to correct for multiple comparisons.

RESULTS

Demographic variables, neuropsychological performances, and volumetric analysis for the entire sample and sub-groups (HOA and MCI) are provided in Table 1. There were no significant differences between groups in terms of age, ethnicity (Caucasian vs. Non-Caucasian), or education. However, there was a significant effect of sex, with a greater proportion of females in the HOA group compared to the MCI group. Results from the MANCOVA (controlling for sex) yielded a significant effect of group for neuropsychological test performances [F(8, 121)=21.024, p<0.001], medial temporal lobe volumes for NQ1 [F(8,122)=2.945, p<0.01] and NQ 2 ICV [F(6,124)=3.03, p<0.01]]. With regards to NQ2 z-scores, independent samples *t*-tests revealed significant differences

between groups in the hippocampus and amygdala, but not in the isthmus and post-cingulate volumes.

Aim 1: Replication using NQ1 and NQ2 (see Tables 2 and 3)

We replicated earlier findings of a positive relationship between DMI performance and hippocampal volume across all participants and in the MCI group; DMI was also negatively correlated with total inferior lateral ventricle across participants. Although IMI and LI were negatively correlated with inferior lateral ventricle across all participants in our previous study [13], neither of these indices were statistically significant in the current study using either NQ1 or NQ2. The AI was significantly and negatively correlated with ILV volume for NQ1 and NQ2 across all participants, but not in the subgroups; this finding was driven by performance on the coding subtest in NQ1 ($r= -.253$, $p=.004$) and NQ2 ($r= -.247$, $p=.004$) rather than digit span (NQ: $r= -.060$, $p=.494$; NQ2: $r= -.072$, $p= .415$).

Regarding the newly included amygdala, we found significant positive relationships between volume and all indices except visuospatial/construction regardless of NQ version in all participants. In the HOA group, amygdala volume was significantly and positively correlated with LI for both NQ versions; this correlation was driven by semantic fluency performance (NQ1: $r=.418$, $p=<.001$; NQ2: $r=.366$, $p=.002$) rather than naming (NQ1: $r= -.055$, $p=.660$; NQ2: $r=.022$, $p=.858$). DMI was significantly and positively correlated with the amygdala in the MCI sample using NQ1, but failed to surpass the FDR correction threshold for NQ2.

Aim 2: Agreement between NQ1 and NQ2

We computed B&A plots to evaluate any systematic difference between NQ versions. First, we calculated the differences in the mean volumes (NQ1 – NQ2; a negative value indicates that NQ2 produced a larger value than NQ1 whereas a positive value indicates a larger value in NQ1 than NQ2) between the NQ versions and established 95% limits of agreement for each of the 11 regions that are common to both versions. We also reviewed a random sample of outliers from the plots and determined that these differences were not the result of manual data entry or NQ processing errors. In order to determine if there was a significant difference between the NQ versions, we conducted a one-sample t-test of the mean difference value for each of the regions for each hemisphere, as well as, total volume (i.e., left + right) for each region. With the exception of the left inferior lateral ventricle and right/total cerebellum, all volumes were significantly different from one another (Table 4). Select plots are provided in Figure 1 and demonstrate that NQ1 provides larger values for the right and left hippocampus and amygdala and NQ2 provides larger values for the right inferior lateral ventricle.

Aim 3: Utility of NQ2 normatively-derived z-scores

There were significant relationships between NQ2 hippocampus z-scores and DMI in the entire sample and in the MCI group when controlling for NQ2 %ICV (Table 5). However, there were no significant relationships between NQ2 %ICV and either memory index when controlling for NQ2 z-scores (Table 6). A similar finding was evident for the left amygdala and IMI in the MCI group. There were no statistically significant findings in the HOA group.

DISCUSSION

The integration of MRI-derived volumetry into clinical practice is limited by a number of practical and methodological barriers, including the ability to apply knowledge of group-level differences to the individual patient. NQ offers a standardized and clinical-friendly method of volumetric analysis that could potentially address several of the aforementioned barriers, including the use of normatively-derived volumetrics. This is the first study, to our knowledge, to independently compare volumes drawn from NQ versions and their relationship with neuropsychological functioning, as well as to evaluate the utility of normatively-adjusted volumetric data in HOA and MCI.

With respect to our first aim, we generally replicated previous findings [13] between key medial temporal lobe regions and memory test performance using NQ1 and NQ2. Compared to HOA, the MCI subgroup exhibited generally smaller medial temporal lobe volumes, which is consistent with the neuropathological changes in MCI compared to normal aging [29]. Of the three included medial temporal volumes, only the hippocampus was related to DMI regardless of NQ version. This finding supports the large body of evidence for the relationship between this structure and memory performance [30–31] and further highlights its specificity in terms of cognitive test performance (at least based on the current set of measures). However, unlike earlier studies that demonstrated a relationship between posterior cingulate volumes in MCI [23], there were no differences the isthmus/post-cingulate volumes in our current study. This raises the possibility that volumetric decline may not become apparent until later in the disease course and/or that methodological differences between the regional definitions in various volumetric approaches may be responsible for these findings.

While the above findings generally supported the role of the MTL in memory (i.e., convergent validity), it is equally important to establish divergent validity. In this respect, it is important to note that the hippocampus was only related to memory test performance, both in the entire sample and in the MCI group. In contrast, the volume of the amygdala appeared to be a rather non-specific marker of cognitive change given its relationships with nearly every RBANS index in the entire sample. The relationship between the amygdala and semantic fluency in the HOA group may be meaningful given the combined findings of poor performance in those with MCI [32] and evidence of an anterior to posterior pattern of MTL atrophy in Alzheimer's disease [33]. Future studies will need to determine the predictive validity of this relationship. The overall patterns of findings are important to consider since memory was consistently related to hippocampal volume in both our current and previous [13] studies. In contrast, relationships between cognition and ILV differed across these studies and NQ versions.

When comparing level of agreement between NQ versions, our findings revealed substantially different medial temporal lobe values despite the common neuroanatomical label. Specifically, most NQ1 volumes were significantly different from NQ2 volumes with the exception of the left inferior lateral ventricle and right/total cerebellum. However, there did not seem to be any particular pattern as a function of version, subcortical vs. neocortical region, or cerebral hemisphere. These findings indicate that NQ1 and NQ2 values are not

interchangeable and warrant caution when attempting to use particular values cutoffs that are based on NQ1 data. We did not determine whether one NQ version is more accurate than the other since this would require manual-tracing or other well-established methods [10, 34] including histopathological correlation. Regardless, medial temporal lobe regions generally maintained relationships with learning and memory performances in both NQ version, as in our earlier report [13]. One exception relative to our earlier report [13] is the lack of significant relationship between inferior lateral ventricle volume and memory test performance.

An especially novel aspect of NQ2 is the extensive normative sample and associated age and sex-based z-scores, which provide the context critical for translating volumetrics to the individual patient level. Our results also indicate that the NQ2 z-scores are not only more clinically applicable than standard ICV values, but they also explain significantly more variance in the relationship between hippocampus and DMI above and beyond standard %ICV correction. Such findings reinforce the use of normatively-derived volumes [35] and suggest these values can be particularly helpful at the individual patient level

While our findings are encouraging, several important limitations should be noted. First, our participants were characterized according to their cognitive phenotype. This is consonant with the Petersen 2004 criteria [15] used at the time of enrollment but fails to establish the presence of characteristic AD biomarkers as suggested by current criteria [36]. While future studies should integrate beta-amyloid and tau in the diagnostic process to ensure a homogeneous disease etiology, we are encouraged by the smaller MTL volumes in our MCI sample since multiple models [4, 37] and empirical evidence [38–39] link reduced MTL volumes with regional AD pathology. Our study is also cross-sectional and limited to healthy older adults and those with MCI, so we cannot comment on the utility of the measures to predict clinical conversion (i.e., from HOA to MCI or MCI to dementia). Future research is needed to evaluate such longitudinal uses as well as how values differ in other MCI subtypes and disease etiologies (e.g., non-amnestic MCI and vascular dementia). However, our current infrastructure will address both of these weaknesses using a new dataset since many of our longitudinally-monitored participants have consented to brain donation (via the Michigan Alzheimer's Disease Center). Our sample was generally matched in terms of age, education, and ethnicity, but there remained a statistically greater number of females compared to males in the HOA group. Therefore, we utilized statistical methods (partial correlations) to control for this effect; however, we acknowledge that this should be further examined in future studies given literature suggesting considerable cognitive variability as a function of sex in those with MCI [40] as well as potential interaction effects between sex and biomarkers [41–42]. Relatedly, our sample was well educated and results may not generalize to individuals with lower educational achievement. We cannot rule out potential confounds related to the recruitment sources for the HOA and MCI groups; however, such factors seem unlikely given the 1) common registry used for recruitment of most of the participants (i.e., Emory ADRC) and 2) use of the same neuropsychological protocol to ensure cognitive phenotype. While the groups were not significantly different in age, it is possible that the trend-level difference contributed to the findings. However, we did not control for age because: 1) the effect size difference in hippocampal volumes was comparable for the NQ2 approaches (Cohen's d for ICV = .61; for z-score = .61), 2) the z-

score volumes and RBANS data were already age corrected, 3) the correlation results remained significant in only the MCI group, which argues against a primary age-effect driving the findings for the entire sample. For exploratory purposes, we reran partial correlations between the hippocampus and RBANS DMI correcting for both age and sex in the entire sample; the significance values were unchanged (all p<.001) and the coefficients were nominally different (NQ1 ICV= .408 – change of .01; NQ2 ICV= .395 – change of .054; NQ2 z-score = .400 – change of .004). Thus, there is no clear evidence to suggest age affected our results. Finally, there are 39 brain regions provided in NQ2 z-scores, so further research is needed to evaluate the utility of these data beyond medial temporal lobe volumes.

Overall, our findings generally validate and extend our earlier work [13]. Although the precise values between NQ ICV version values differed, medial temporal lobe volumes remained significantly related to memory test performance, especially in those with MCI. The z-score volumes are inherently useful from an individual patient standpoint and appear to be more sensitive to memory test performance, at least in NQ2. Taken as a whole, our findings suggest that NQ is a viable tool for the evaluation and interpretation of medial temporal lobe volume in older adults.

Acknowledgments

References

1. Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, Smith GE, Ivnik RJ, Kokmen E. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology. 1997; 49:786–794. [PubMed: 9305341]

2. Braskie MN, Thompson PM. A focus on structural brain imaging in the Alzheimer's disease neuroimaging initiative. Biol Psychiatry. 2014; 75:527–533. [PubMed: 24367935]

3. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E, Morris JC, Petersen RC, Saykin AJ, Schmidt ME, Shaw L, Shen L, Siuciak JA, Soares H, Toga AW, Trojanowski JQ. Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's disease neuroimaging initiative: A review of papers published since its inception. Alzheimers Dement. 2013; 9:e111–e194. [PubMed: 23932184]

4. Jack CR Jr, Holtzman DM. Biomarker modeling of Alzheimer's disease. Neuron. 2013; 18:1347–1358.

5. Brewer JB. Fully-automated volumetric MRI with normative ranges: Translation to clinical practice. Behav Neurol. 2009; 21:21–28. [PubMed: 19847042]

6. Hampstead BM, Brown GS. Using neuroimaging to inform clinical practice for the diagnosis and treatment of mild cognitive impairment. Clin Geriatr Med. 2013; 29:829–845. [PubMed: 24094299]

7. Luo, W., Airriess, C., Albright, J. The NeuroQuant normative database: Comparing individual brain structures. CorTechs Labs, Inc; 2015. http://www.cortechslabs.com/whitepapers [Accessed 16 January 2017]

8. Brewer JB, Magda S, Airriess C, Smith ME. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. AM J Neuroradiol. 2009; 30:578–580. [PubMed: 19112065]

9. Ross DE, Graham TJ, Ochs AL. Review of the evidence supporting the medical and legal use of NeuroQuant in patients with traumatic brain injury. Psychol Inj and Law. 2012; 6:75–80.

10. Ochs AL, Ross DE, Zannoni MD, Abildskov TJ, Bigler ED. Alzheimer's Disease Neuroimaging Initiative. Comparison of automated brain volume measures obtained with Neuroquant and Free Surfer. J Neuroimaging. 2015; 25:721–727. [PubMed: 25727700]

11. Kovacevic S, Rafii MS, Brewer JB. Alzheimer's Disease Neuroimaging Initiative. High-throughhippocampusut, fully automated volumetry for prediction of MMSE and CDR decline in mild cognitive impairment. Alzheimer Dis Assoc Disord. 2009; 23:139–45. [PubMed: 19474571]

12. Gold JJ, Squire LR. Quantifying medial temporal lobe damage in memory-impaired patients. Hippocampus. 2005; 15:79–85. [PubMed: 15390163]

13. England HB, Gillis MM, Hampstead BM. RBANS memory indices are related to medial temporal lobe volumetrics in healthy older adults and those with mild cognitive impairment. Arch Clin Neuropsychol. 2014; 29:322–328. [PubMed: 24709384]

14. White, N., Magda, S., Airriess, C., Albright, J. Dynamic atlas: The new personalized segmentation approach. Cortechs Labs, Inc; 2015. http://www.cortechslabs.com/whitepapers [Accessed 16 January 2017]

15. Peterson RC. Mild cognitive impairment as a diagnostic entity. J Intern Med. 2004; 256:183–194. [PubMed: 15324362]

16. Randolph C, Tierney MC, Mohr E, Chase TN. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary clinical validity. J Clin Exp Neuropsychol. 1998; 20:310–319. [PubMed: 9845158]

17. Reitan, RM., Wolfson, D. The Halstead–Reitan Neuropsychological Test Battery: Therapy and clinical interpretation. Tucson, AZ: Neuropsychological Press; 1985.

18. Sheikh, JA., Yesavage, JA. Geriatric Depression Scale (GDS): Recent findings and development of a shorter version. In: Brink, TL., editor. Clinical Gerontology: A guide to assessment and intervention. Howarth Press; New York: 1986. p. 165-174.

19. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol. 1982; 37:323–329. [PubMed: 7069156]

20. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975; 12:189–198. [PubMed: 1202204]

21. Csukly G, Siraly E, Fodor Z, Horvath A, Salacz P, Hidasi Z, Csibri E, Rudas G, Szabo A. The differentiation of amnestic type MCI from the non-amnestic types by structural MRI. Front Aging Neurosci. 2016; 8 eCollection 2016.

22. Adams, S. Personal communication. 2017 Jan 16.

23. Han Y, Lui S, Kuang W, Lang Q, Zou L, Jia J. Anatomical and functional deficits in patients with amnestic mild cognitive impairment. PLoS One. 2012; 7:e28664. [PubMed: 22319555]

24. Chetelat G, Landeau B, Eustache F, Mezenge F, Viader F, de la Sayette V, Desgranges B, Baron JC. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. Neuroimage. 2005; 27:934–946. [PubMed: 15979341]

25. Kendall, MG., Stuart, A. The advanced theory of statistics. London: C. Griffin; 1958.

26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310. [PubMed: 2868172]

27. Giavarina D. Understanding Bland Altman analysis. Biochem Med. 2015; 25:141–151.

28. Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage. 2009; 44:62–70. [PubMed: 18603449]

29. Driscoll I, Davatzikos C, An Y, Wu X, Shen D, Kraut M, Resnick SM. Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. Neurology. 2009; 72:1906–1913. [PubMed: 19487648]

30. Mungas D, Harvey D, Reed BR, Jaqust WJ, DeCarli C, Beckett L, Mack WJ, Kramer JH, Weiner MW, Schuff N, Chui HC. Longitudinal volumetric MRI change and rate of cognitive decline. Neurology. 2005; 65:565–571. [PubMed: 16116117]

31. Van de Pol LA, Van der Flier WM, Korf ES, Fox NC, Barkhof F, Scheltens P. Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment. Neurology. 2007; 69:1491–1497. [PubMed: 17923611]

32. Teng E, Leone-Friedman J, Lee GJ, Woo S, Apostolova LG, Harrell S, Ringman JM, Lu PH. Similar verbal fluency patterns in amnestic mild cognitive impairment and Alzheimer's disease. Arch Clin Neuropsychol. 2013; 28:400–410. [PubMed: 23752677]

33. Whitwell JL. Progression of atrophy in Alzheimer's disease and related disorders. Neurotox Res. 2010; 18:339–346. [PubMed: 20352396]

34. Lyden H, Gimbel SI, Del Piero L, Tsai AB, Sachs ME, Kaplan JT, Margolin G, Saxbe D. Associations between family adversity and brain volume in adolescence: Manual vs. automated brain segmentation yields different results. Front Neurosci. 2016; 10 eCollection 2016.

35. Merkel B, Steward C, Vivash L, Malpas CB, Phal P, Moffat BA, Cox KL, Ellis KA, Ames DJ, Cyarto EV, Lai MM, Sharman MJ, Szoeke C, Masters CL, Lautenschlager NT, Desmond P. Semi-automated hippocampal segmentation in people with cognitive impairment using an age appropriate template for registration. J Magn Reson Imaging. 2015; 42:1631–1638. [PubMed: 26140584]

36. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jaqust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011; 7:270–290. [PubMed: 21514249]

37. Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, Schott JM, Alexander DC. Alzheimer's Disease Neuroimaging Initiative. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain. 2014; 137:2564–2577. [PubMed: 25012224]

38. Henneman WJ, Vrenken H, Barnes J, Sluimer IC, Verwey NA, Blankenstein MA, Klein M, Fox NC, Scheltens P, Barkhof F, Van der Flier WM. Baseline CSF p-tau levels independently predict progression of hippocampal atrophy in Alzheimer's disease. Neurology. 2009; 73:935–940. [PubMed: 19770469]

39. Li JQ, Tan L, Wang HF, Tan MS, Tan L, Xu W, Zhao QF, Wang J, Jiang T, Yu JT. Risk factors for predicting progression from mild cognitive impairment to Alzheimer's disease: A systematic review and meta-analysis of cohort studies. J Neurol Neurosurg Psychiatry. 2016; 87:476–484. [PubMed: 26001840]

40. Au B, Dale-McGrath S, Tierney MC. Sex differences in the prevalence and incidence of mild cognitive impairment: A meta-analysis. Ageing Res Rev. 2017; 35:176–199. [PubMed: 27771474]

41. Fleisher A, Grundman M, Jack CR Jr, Petersen RC, Taylor C, Kim HT, Schiller DH, Bagwell V, Sencakova D, Weiner MF, DeCarli C, DeKosky ST, van Dyck CH, Thal LJ. Alzheimer's Disease Cooperative Study. Sex, apolipoprotein E epsilon 4 status, and hippocampal volume in mild cognitive impairment. Arch Neurol. 2005; 62:953–957. [PubMed: 15956166]

42. Sundermann EE, Biegon A, Rubin LH, Lipton RB, Mowrey W, Landau S, Maki PM. Alzheimer's Disease Cooperative Study. Better verbal memory in women than men in MCI despite similar levels of hippocampal atrophy. Neurology. 2016; 86:1368–1376. [PubMed: 26984945]
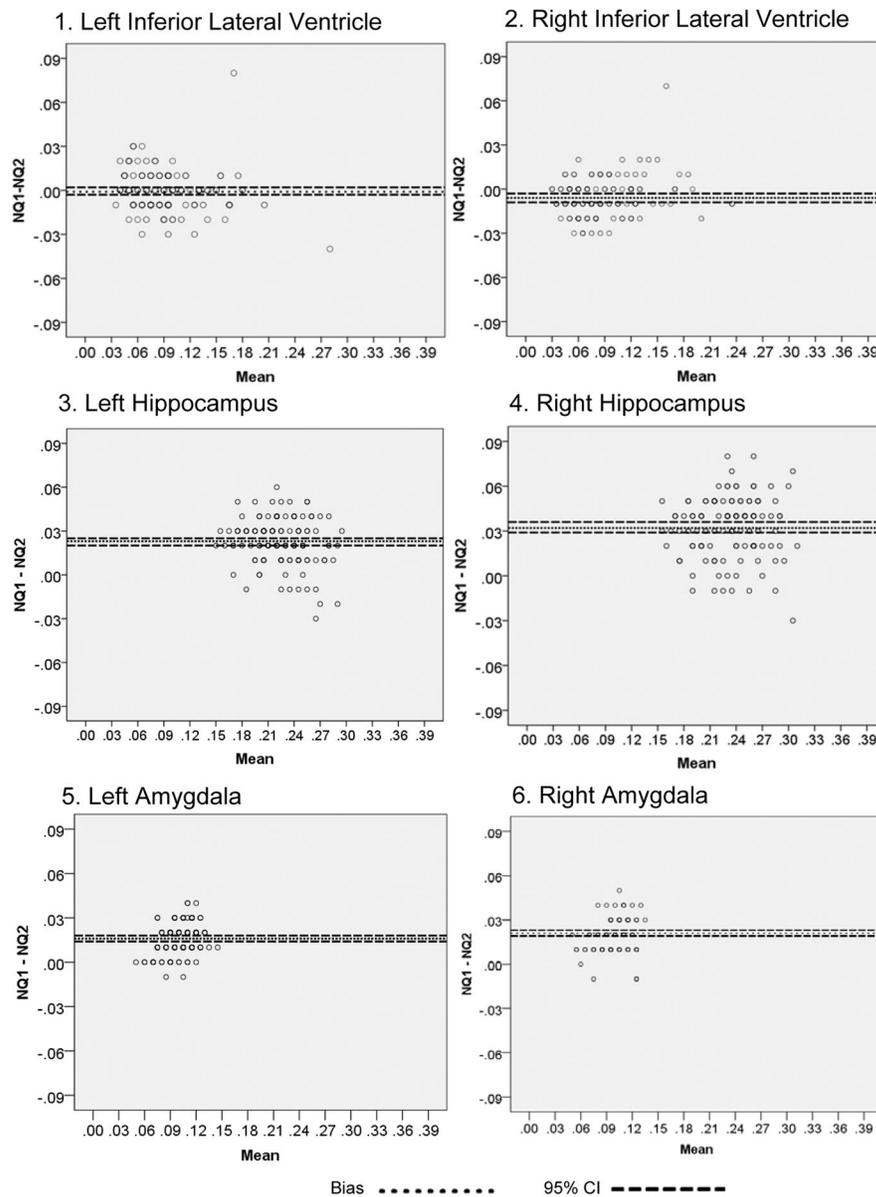
Figure 1.
Bland-Altman plots of representative medial temporal lobe volumes with 95% confidence intervals (dashed lines). Positive values indicate larger NQ1 values while negative indicate larger NQ2.

Table 1

Descriptive characteristics for all participants as well as the healthy older adults (HOA) and mild cognitive impairment (MCI) subgroups.

	Total (n=132) M(SD)	HOA (n=67) M(SD)	MCI (n=65) M(SD)	Statistic (HOA vs. MCI)	p-value
Age (years)	71.63 (7.20)	70.61 (6.31)	72.68 (7.93)	t(122.109)=1.653	0.10
Education (years)	16.46 (2.35)	16.60 (2.04)	16.32 (2.65)	t(120.235)=0.665	0.51
Ethnicity (frequency)				$x2$=0.476 (1)	0.56
Caucasian	95	50	45		
African American	32	15	17		
Latino	5	2	3		
Sex (frequency)				$x2$=6.94 (1)	**0.008**
Male	52	19	33		
Female	80	48	32		
MMSE (raw score)	28.35 (1.94)	29.24 (1.24)	27.43 (2.10)	*F(8,121)* *31.693*	**<.001**
Trails A (z-score)	−0.14 (1.07)	0.06 (0.86)	−0.34 (1.22)	*2.684*	0.104
Trails B (z-score)	−0.06 (1.00)	0.13 (0.93)	−0.26 (1.04)	*4.255*	0.041
RBANS Indices (z-score)					
Immediate Memory (IMI)	−0.13 (1.17)	0.64 (0.72)	−0.93 (0.99)	97.985	**<.001**
Visuospatial/Constructional (VI)	−0.03 (1.01)	0.21 (0.93)	−0.28 (1.04)	8.503	**0.004**
Language (LI)	−0.18 (0.91)	0.24 (0.73)	−0.61 (0.89)	28.363	**<.001**
Attention (AI)	0.15 (0.94)	0.45 (0.83)	−0.17 (0.95)	17.526	**<.001**
Delayed Memory (DMI)	−0.47 (1.33)	0.47 (0.54)	−1.44 (1.20)	130.068	**<.001**
NQI: Brain Volumes (% ICV)				*F(8,122)*	
Left inferior lateral ventricle	0.09 (0.04)	0.08 (0.03)	0.10 (0.04)	6.462	**0.012**
Right inferior lateral ventricle	0.09 (0.04)	0.07 (0.03)	0.10 (0.05)	9.918	**0.002**
Total inferior lateral ventricle	0.18 (0.08)	0.15 (0.06)	0.20 (0.09)	8.744	**0.004**
Left hippocampus	0.24 (0.03)	0.25 (0.03)	0.22 (0.03)	12.100	**0.001**
Right hippocampus	0.25 (0.04)	0.26 (0.03)	0.24 (0.04)	8.784	**0.004**
Total hippocampus	0.48 (0.07)	0.51 (0.05)	0.46 (0.07)	14.505	**<.001**
Left amygdala	0.11 (0.02)	0.11 (0.02)	0.10 (0.02)	9.116	**0.003**
Right amygdala	0.11 (0.02)	0.12 (0.01)	0.11 (0.02)	11.167	**0.001**

	Total (n=132) M(SD)	HOA (n=67) M(SD)	MCI (n=65) M(SD)	Statistic (HOA vs. MCI)	p-value
Total amygdala	0.22 (0.04)	0.23 (0.03)	0.21 (0.04)	11.742	**0.001**
NQ2: Brain Volumes (% ICV)				$F(6,124)$	
Left inferior lateral ventricle	0.09 (0.04)	0.08 (0.03)	0.11 (0.05)	6.882	**0.010**
Right inferior lateral ventricle	0.09 (0.04)	0.08 (0.03)	0.10 (0.05)	10.984	**0.001**
Total inferior lateral ventricle	0.18 (0.08)	0.16 (0.06)	0.21 (0.09)	9.481	**0.003**
Left hippocampus	0.21 (0.03)	0.22 (0.03)	0.20 (0.03)	8.442	**0.004**
Right hippocampus	0.22 (0.04)	0.23 (0.03)	0.21 (0.04)	8.841	**0.004**
Total hippocampus	0.43 (0.07)	0.45 (0.06)	0.41 (0.07)	9.377	**0.003**
Left amygdala	0.09 (0.02)	0.10 (0.01)	0.09 (0.02)	9.118	**0.003**
Right amygdala	0.09 (0.02)	0.10 (0.01)	0.09 (0.02)	12.590	**0.001**
Total amygdala	0.18 (0.03)	0.19 (0.02)	0.17 (0.03)	12.464	**0.001**
NQ2: Brain Volumes (z-scores)					
Left hippocampus	−0.11 (1.03)	0.19 (0.97)	−0.42 (1.00)	$t(130)=3.544$	**0.001**
Right hippocampus	−0.11 (1.08)	0.18 (1.00)	−0.40 (1.08)	$t(130)=3.204$	**0.002**
Hippocampus total *	−0.11 (1.00)	0.18 (0.93)	−0.41 (0.99)	$t(130)=3.547$	**0.001**
Left amygdala	−0.29 (1.06)	−0.02 (1.03)	−0.56 (1.04)	$t(130)=2.983$	**0.003**
Right amygdala	−0.15 (1.02)	0.16 (0.91)	−0.47 (1.03)	$t(130)=3.735$	**<.001**
Amygdala total *	−0.22 (0.98)	0.07 (0.91)	−0.52 (0.96)	$t(130)=3.586$	**<.001**
Left isthmus and post-cingulate	0.00 (0.96)	0.10 (0.91)	−0.11 (1.00)	$t(130)=1.256$	0.211
Right isthmus and post-cingulate	0.02 (1.04)	0.06 (1.07)	−0.01 (1.02)	$t(130)=0.373$	0.710
Isthmus and post-cingulate total *	0.01 (0.94)	0.08 (0.93)	−0.06 (0.96)	$t(130)=0.846$	0.399

Note:

*
average z-score; Trails B sample size is n=131 for total sample (n=64 for MCI).

Bolded values were below the FDR correction threshold of p = 0.026 (#tests=39)

Table 2

Partial correlations (*p*-values) between RBANS indices and NQ1 (% total ICV) total inferior lateral ventricle (ILV), hippocampus (HP) and amygdala (AMG) volumes while controlling for sex

	All Participants (N=132)			HOA (N=67)			MCI (N=65)		
	ILV	HP	AMG	ILV	HP	AMG	ILV	HP	AMG
IMI	−.136 (.121)	.104 (.239)	**.228 (.009)**	.016 (.898)	−.203 (.102)	.025 (.843)	.035 (.782)	−.111 (.384)	.090 (.480)
VI	−.067 (.447)	.020 (.818)	.034 (.698)	−.017 (.892)	−.053 (.670)	−.030 (.809)	−.015 (.904)	−.074 (.561)	−.029 (.820)
LI	−.092 (.295)	.109 (.217)	**.210 (.016)**	−.198 (.111)	.068 (.589)	**.362 (.003)**	.144 (.255)	−.104 (.416)	−.048 (.707)
AI	**−.196 (.025)**	.182 (.037)	**.220 (.012)**	−.140 (.262)	.135 (.281)	.117 (.351)	−.123 (.331)	.040 (.753)	.164 (.195)
DMI	**−.270 (.002)**	**.398 (<.001)**	**.363 (<.001)**	−.037 (.771)	.075 (.549)	.088 (.480)	−.179 (.157)	**.346 (.005)**	**.302 (.015)**

Bolded values were below the FDR correction threshold of p = 0.026 (#tests=45). IMI= Immediate Memory Index; VI= Visuospatial/Constructional Index; LI= Language Index; AI= Attention Index; DMI = Delayed Memory Index.

Table 3

Partial correlations (*p*-values) between RBANS indices and NQ2 (% total ICV) total inferior lateral ventricle, hippocampus, and amygdala volumes while controlling for sex

	All Participants (N=132)			HOA (N=67)			MCI (N=65)		
	ILV	**HP**	**AMG**	**ILV**	**HP**	**AMG**	**ILV**	**HP**	**AMG**
IMI	−.101	.162	**.296**	.034	−.117	.137	.121	.072	.167
	(.252)	(.064)	(**.001**)	(.785)	(.350)	(.272)	(.340)	(.574)	(.187)
VI	−.061	.033	.091	−.003	.062	.116	.001	−.112	−.031
	(.491)	(.711)	(.301)	(.983)	(.621)	(.353)	(.994)	(.377)	(.806)
LI	−.114	.192	**.209**	−.220	.225	**.297**	.123	−.010	−.028
	(.196)	(.028)	(**.017**)	(.076)	(.070)	(**.015**)	(.331)	(.940)	(.825)
AI	**−.209**	.142	**.203**	−.167	.116	.073	−.118	.013	.155
	(**.016**)	(.106)	(**.020**)	(.181)	(.355)	(.562)	(.353)	(.921)	(.220)
DMI	**−.211**	**.341**	**.373**	−.033	.087	.223	−.065	**.319**	.270
	(**.016**)	(**<.001**)	(**<.001**)	(.792)	(.487)	(.072)	(.612)	(**.010**)	(.031)

Bolded values were below the FDR correction threshold of p = 0.026 (#tests=45). IMI= Immediate Memory Index; VI= Visuospatial/Constructional Index; LI= Language Index; AI= Attention Index; DMI = Delayed Memory Index.

Table 4

Mean Difference between NQ1 and NQ2 (ICV) for Bland-Altman Plots

Region (%ICV)	Mean Difference	t(131)	p-value	95% CI Lower	95% CI Upper
Left Forebrain Parenchyma	−0.565	22.510	<.001	−0.614	−0.520
Right Forebrain Parenchyma	−0.499	16.699	<.001	−0.558	−0.440
Total Forebrain Parenchyma	−1.064	20.311	<.001	−1.167	−0.960
Left Cortical Gray Matter	−0.181	6.502	<.001	−0.236	−0.126
Right Cortical Gray	−0.129	3.735	<.001	−0.198	−0.061
Total Cortical Grey Matter	−0.310	5.462	<.001	−0.423	−0.198
Left Lateral Ventricle	−0.037	2.480	0.014	−0.067	−0.008
Right Lateral Ventricle	−0.085	7.898	<.001	−0.106	−0.063
Total Lateral Ventricle	−0.122	4.828	<.001	−0.171	−0.072
Left Inferior Lateral Ventricle	−0.001	0.539	0.591	−0.003	0.002
Right Inferior Lateral Ventricle	−0.005	4.702	<.001	−0.008	−0.003
Total Inferior Lateral Ventricle	−0.006	2.797	0.006	−0.011	−0.002
Left Hippocampus	0.023	16.454	<.001	0.020	0.025
Right Hippocampus	0.032	18.913	<.001	0.029	0.036
Total Hippocampus	0.053	13.831	<.001	0.045	0.060
Left Amygdala	0.016	17.591	<.001	0.014	0.018
Right Amygdala	0.021	23.119	<.001	0.019	0.023
Total Amygdala	0.037	23.890	<.001	0.034	0.040
Left Caudate	0.070	48.332	<.001	0.067	0.073
Right Caudate	0.071	43.400	<.001	0.068	0.075
Total Caudate	0.141	51.510	<.001	0.135	0.146
Left Putamen	−0.032	14.304	<.001	−0.036	−0.027
Right Putamen	−0.031	12.654	<.001	−0.036	−0.026
Total Putamen	−0.063	14.433	<.001	−0.072	−0.054
Left Pallidum	0.018	13.105	<.001	0.015	0.020
Right Pallidum	0.017	12.349	<.001	0.014	0.020
Total Pallidum	0.035	13.461	<.001	0.030	0.040
Left Thalamus	0.084	30.202	<.001	0.079	0.089

Region (%ICV)	Mean Difference	t(131)	p-value	95% CI	
				Lower	Upper
Right Thalamus	0.098	33.608	<.001	0.093	0.104
Total Thalamus	0.182	40.197	<.001	0.173	0.191
Left Cerebellum	0.024	3.826	<.001	0.012	0.037
Right Cerebellum	0.068	1.458	0.147	−0.024	0.159
Total Cerebellum	0.091	1.931	0.056	0.002	0.185

Table 5

Partial correlations (*p*-values) between NQ2 z-score and memory test performance while controlling for and NQ2 ICV of hippocampal (HP) and amygdala (AMG) volumes

	Left HP	Right HP	Average HP	Left AMG	Right AMG	Average AMG
All participants (N=132)						
IMI (z-score)	.056 (.527)	.077 (.380)	.079 (.369)	.095 (.281)	.088 (.319)	.099 (.262)
DMI (z-score)	.192 (.028)	.174 (.046)	**.205 (.019)**	.053 (.545)	.062 (.481)	.060 (.500)
HOA (N=67)						
	Left HP	Right HP	Average HP	Left AMG	Right AMG	Average AMG
IMI (z-score)	−.119 (.339)	.052 (.677)	−.026 (.834)	.066 (.598)	.133 (.287)	.104 (.407)
DMI (z-score)	.061 (.629)	.226 (.068)	.164 (.187)	.136 (.275)	.152 (.223)	.153 (.221)
MCI (N=65)						
	Left HP	Right HP	Average HP	Left AMG	Right AMG	Average AMG
IMI (z-score)	.179 (.157)	.182 (.149)	.196 (.120)	**.312 (.012)**	.100 (.432)	.236 (.060)
DMI (z-score)	**.351 (.005)**	**.321 (.010)**	**.364 (.003)**	.190 (.133)	.102 (.424)	.169 (.181)

Bolded values were below the FDR correction threshold of p = 0.026 (#tests=36)

Table 6

Partial correlations (*p*-values) between NQ2 ICV and memory test performance while controlling for and NQ2 z-score of hippocampal (HP) and amygdala (AMG)volumes

	All participants (N=132)					
	Left HP	**Right HP**	**Average HP**	**Left AMG**	**Right AMG**	**Average AMG**
IMI (z-score)	.052 (.557)	.024 (.788)	.029 (.745)	.046 (.601)	.044 (.620)	.040 (.652)
DMI (z-score)	−.018 (.837)	.027 (.756)	−.011 (.904)	.111 (.208)	.098 (.265)	.106 (.229)
	HOA (N=67)					
	Left HP	Right HP	Average HP	Left AMG	Right AMG	Average AMG
IMI (z-score)	.095 (.450)	−.087 (.488)	−.005 (.969)	−.013 (.918)	−.054 (.665)	−.041 (.744)
DMI (z-score)	−.016 (.896)	−.148 (.237)	−.100 (.422)	−.015 (.908)	−.080 (.524)	−.056 (.656)
	MCI (N=65)					
	Left HP	Right HP	Average HP	Left AMG	Right AMG	Average AMG
IMI (z-score)	−.119 (.348)	−.159 (.208)	−.155 (.221)	−.210 (.095)	−.072 (.571)	−.168 (.183)
DMI (z-score)	−.203 (.107)	−.169 (.183)	−.212 (.093)	−.073 (.569)	−.022 (.862)	−.068 (.593)

Bolded values were below the FDR correction threshold of p = 0.026 (#tests=36)