



Published in final edited form as:

Ann Neurol. 2016 December ; 80(6): 846–857. doi:10.1002/ana.24792.

Episodic Memory of Odors Stratifies Alzheimer Biomarkers in Normal Elderly

Alefiya Dhillal Albers, PhD^{1,2}, Josephine Asafu-Adjei, PhD³, Mary K. Delaney¹, Kathleen E. Kelly, BS¹, Teresa Gomez-Isla, MD, PhD¹, Deborah Blacker, MD, ScD⁴, Keith A. Johnson, MD⁵, Reisa A. Sperling, MD¹, Bradley T. Hyman, MD, PhD¹, Rebecca A. Betensky, PhD⁶, Lloyd Hastings, PhD⁷, and Mark W. Albers, MD, PhD¹

¹Department of Neurology, Massachusetts General Hospital, Boston, MA

²Department of Psychology, Endicott College, Beverly MA

³Department of Biostatistics, School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁴Department of Psychiatry, Massachusetts General Hospital, Boston, MA

⁵Department of Radiology, Massachusetts General Hospital, Boston, MA

⁶Department of Biostatistics, Harvard School of Public Health, Boston, MA

⁷Osmic Enterprises, Cincinnati, OH

Abstract

Objective—To relate a novel test of identifying and recalling odor percepts to biomarkers of Alzheimer’s Disease (AD) in well-characterized elderly individuals, ranging from cognitively normal to demented.

Methods—183 participants (cognitively normal: n=70, subjective cognitive concerns: n=74, Mild Cognitive Impairment (MCI): n=29, AD dementia: n=10) were administered novel olfactory tests: the Odor Percept Identification (OPID) and the Percepts of Odor Episodic Memory (POEM) tests. Univariate cross-sectional analyses of performance across diagnoses; logistic regression modeling including covariates of age, gender, education, *APOE* genotype, and neuropsychological test scores; and linear mixed modeling of longitudinal cognitive scores were performed. Amyloid deposition and MRI volumetrics were analyzed in a subset of participants.

Results—Accuracy of identification and episodic memory of odor percepts differed significantly across diagnosis and age, with progressively worse performance across degrees of impairment.

Corresponding author: Mark W. Albers, MD, PhD, MassGeneral Institute of Neurodegenerative Disease, 114 16th St. #2003, Boston, MA 02129, phone and fax: T: 617-724-7401, F: 617-724-1480, albers.mark@mgh.harvard.edu.

Author Contributions:

A.D.A., J.A.-A., B.T.H., R.A.B., L.H., and M.W.A. conceptualized and designed the study. A.D.A., J.A.-A., M.K.D., K.E.K., T.G.I., D.B., K.A.J., R.A.S., and R.A.B. acquired and analyzed data. A.D.A., M.K.D., D.B., and M.W.A. drafted text and prepared figures.

Potential Conflicts of Interest:

Dr. Hastings is the founder, owns stock, and receives a salary for serving as President of Osmic Enterprises, Inc., which markets the OPID and POEM tests. Dr. M. Albers receives consulting fees from International Flavors and Fragrances. All other authors do not have relationships with commercial enterprises that are of direct relevance to the current research.

Among the participants who were cognitively normal or had subjective cognitive concerns, poorer than expected performance on the POEM test (based on the same individual's performance on the OPID and odor discrimination tests) was associated with higher frequencies of the *APOE* ϵ 4 allele, thinner entorhinal cortices, and worse longitudinal trajectory of Logical Memory scores.

Interpretation—Selective impairment of episodic memory of odor percepts, relative to identification and discrimination of odor percepts revealed by this novel POEM battery, is associated with biomarkers of AD in a well-characterized pre-MCI population. These affordable, non-invasive olfactory tests offer potential to identify clinically normal individuals who have greater likelihood of future cognitive decline.

INTRODUCTION

The pathophysiology of AD precedes the onset of cognitive symptoms by a decade or more (1). The validation of indices to identify cognitively healthy individuals at risk for developing the progressive memory symptoms of AD, and to follow these individuals over time is essential for conducting therapeutic trials in this preclinical phase (2). Biomarkers that detect evidence of pathology and are verified for the dementia and MCI phases of AD, e.g., volumetric MRI (3), PET imaging (4) and CSF (5), are under examination in preclinical populations. However, their cost and/or invasiveness may serve as an obstacle to their implementation as a primary screen in clinical trials or in the general population.

Functional assessments, including neuropsychological testing (6), that probe the integrity of vulnerable brain regions may offer sensitivity that is independent of the heterogeneous molecular mechanisms of AD. These assessments may also be sensitive to neurodegeneration, which can occur very early in the preclinical period (7). The regions of the brain that process olfactory input, including the entorhinal cortex and the olfactory bulb, are vulnerable to AD pathology, even in asymptomatic individuals (8). Impaired smell identification is a consistent finding in patients in the dementia and prodromal stages of AD, and has recently been reported in the preclinical stages of AD as well (9). However, such smell identification deficits revealed by traditional tests like the University of Pennsylvania Smell Identification Test (UPSIT) (10) lack specificity to AD as olfactory deficits are also associated with Parkinson's disease (11) as well as aging (12) and end organ perturbations.

To address these limitations, we developed the novel Odor Percept Identification (OPID) and the Percepts of Odor Episodic Memory (POEM) tests. We hypothesized that a task incorporating working memory and episodic memory of odor percepts would improve the efficacy of olfaction as a biomarker of preclinical AD, since performance on these tasks would be expected to reflect the integrity of both the olfactory and memory neural systems (11). An odor percept is the mental impression of an experienced smell isolated from other contextual information. In the UPSIT, response choices are viewed by the participant *before* sampling the test odor. In our ten-item and twenty-item OPID tests, the participant has no visual or semantic information to contextualize the odor that s/he is sniffing. The working memory delay component is introduced into the identification tests by having participants answer a Yes/No question between the odor delivery and the presentation of odor identification choices. In the OPID-10, the Yes/No question is whether the participant finds

the odor familiar; in the OPID-20, the Yes/No question is whether the presented odor was included among the odors presented in the earlier OPID-10. These responses, embedded within the OPID-20, constitute the POEM test. Here we report the validation of this novel olfactory battery administered to a well characterized cohort of elderly individuals representing the continuum of AD and that selective loss of episodic memory of odor percepts identifies clinically normal elderly with an enrichment in other validated biomarkers of AD.

METHODS

Participants

We recruited 183 community-dwelling participants over the age 65 who were recruited from the Longitudinal Cohort (LC) of the Massachusetts Alzheimer's Disease Research Center (B.T.H., P.I.) (13) or from the MGH Memory Disorders Clinic (5 AD patients). Fourteen of these participants were also enrolled in the Harvard Aging Brain Study (R.A.S., P.I.) (4). Our participants were diagnosed at the time of olfactory testing as cognitively normal (CN) (n=70), subjective cognitive concerns (SCC, with concerns about their cognition but had normal cognitive testing) (n=74), or amnesic or non-amnesic mild cognitive impairment (MCI) (n=29) and possible or probable AD (n=10) (Table 1). All participants underwent informed consent before participation. The research protocol was approved by the Institutional Review Board of Partners Healthcare.

Eligibility Criteria

Criteria for enrollment included good general health without any unstable medical diagnoses or active psychiatric disorders, and a study partner who could provide collateral information about the participant's cognition and daily functioning. At the initial visit, participants were 54 years of age, had a Geriatric Depression Scale (14) long form score ≤ 10 , and had a Modified Hachinski Ischemic Score ≥ 4 (15). Participants in our sample were recruited to the ADRC, on average, 5.4 years, prior to their first visit for the olfactory study.

Olfactory screening

Participants (and study partners, when appropriate) completed a questionnaire that surveyed medical factors that may confound olfactory function (septal deviation, difficulty breathing through one side of nose, history of radiation or chemotherapy, history of nasal surgery) (16). Participants were excluded from the study if they presented with respiratory distress or an active sinus / upper respiratory infection at the time of testing, known nasal polyps, a history of anaphylaxis to nuts, traumatic or congenital anosmia, current or recent (past 6 months) alcohol or substance dependence, or pregnancy. No participants were excluded from the study based upon these criteria.

Olfactometer

The olfactory battery was administered in a quiet room with a research assistant present to provide instructions and assist with entering response choices into the laptop driving the portable OLFACT olfactometer (Osmic Enterprises, Cincinnati, OH). The olfactometer has a dedicated line for the delivery of each odor to prevent cross-contamination. Each response

choice and reaction time are recorded by the software program that triggers the olfactometer. For each test, participants were seated next to the olfactometer, given test instructions, and underwent a trial run of odor presentation and response choice to acclimate to the timing and tasks of the test. All collected data were imported and stored in a customized MySQL database.

Olfactory Evaluation

Three olfactory tests were administered, in the same order:

1. The OPID-10 (Figure 1A) employs 10 odors found to be predictive for conversion from MCI to AD, i.e., menthol, clove, leather, strawberry, lilac, pineapple, smoke, soap, grape and lemon (16). Participants were visually cued to get ready to sniff and then presented an odor for 2 s. Subsequently, they were asked “Is this odor familiar to you?” (Yes/No), and then were visually presented four words and asked to select the word that best identified the presented odor.
 - The Odor Awareness Scale (17) (OAS), which provides a measure of how attentive participants are to their olfactory sense and how influenced they are behaviorally and emotionally by their olfactory perceptions, was administered between tests 1 and 2 (and served as a 10-minute delay).
2. A 20-item test (10 odors presented in the OPID-10 and 10 novel odors) that constitute the POEM test and OPID-20 tests (Figure 1B). The protocol is similar to the OPID-10 except the question following the presentation of the odor is “Did you smell this odor in the previous test?” (Yes/No).
3. The odor discrimination test (OD) (Figure 1C) employs 12 trials where 2 odors are presented consecutively, and participants are asked if the odors were the “Same” or “Different”.

Clinical / Neuropsychological / Neuroimaging Evaluations

All participants underwent comprehensive evaluations consisting of a medical history, neuropsychological battery (18), and neurological examination prior to olfactory testing. Visual naming, to control for any aphasic contribution, was assessed with the 30-item Boston Naming Test (BNT) performed at the time of olfactory testing. After an interview with the participants and informants, clinicians rated each participant on the Clinical Dementia Rating-Sum of Boxes (CDR-SOB) in the domains of memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care (19). Each subject was tested at each visit with the MMSE, Logical Memory II (LMII), Digit-Symbol Decoding (DSD) and delayed recall on the California Verbal Learning Test (CVLT). MRI and PiB PET data collected from participants participating in concurrent studies were assembled; the timespan between the neuroimaging evaluations and olfactory testing averaged 2.1 years (5.1 years before to 0.8 years after olfactory testing; imaging preceded olfactory testing for 88% of subjects) for PiB and 1.4 years (5.0 years before to 0.8

years after olfactory testing; imaging preceded olfactory testing for 84% of subjects) for MRI scans.

Structural MRI acquisition and processing

MRI scanning was completed on 49 participants as previously described (3). Briefly, in a Siemens TIM Trio 3T System with a 12-channel head coil, structural T1-weighted volumetric magnetization-prepared, rapid acquisition gradient echo scans were collected (TR/TE/TI=6400/2.8/900ms, flip angle=8°, 1×1×1.2mm resolution). Region of interest labeling was performed using FreeSurfer v5.1 (<http://surfer.nmr.mgh.harvard.edu/>). Ex vivo measurements of entorhinal cortex (EC) thickness were averaged across hemispheres to yield mean bilateral EC thickness values. Hippocampal volume (HV) was averaged across hemispheres and divided by intracranial volume (ICV) to adjust for variance in head size across patients.

PiB-PET acquisition and processing

PiB-PET scanning was completed in 41 participants on a Siemens ECAT EXACT HR+ PET scanner. ¹¹C-PiB was synthesized using a previously published protocol (20). After injection of 8.5 – 15 mCi PiB, 60 min. of dynamic data were acquired in 3D acquisition mode. PiB-PET data was reconstructed using filtered back projection with weighted attenuation. The Logan graphical analysis method with cerebellar cortex as the reference tissue input function was used to evaluate specific PiB retention expressed as the distribution volume ratio (DVR). PiB DVR was calculated for an aggregate of cortical regions that typically have elevated PiB retention in AD dementia. Binary PiB status was assigned with a cut-off value of 1.20 DVR for the aggregate described above (DVR ≥ 1.20, PiB-positive; DVR < 1.20, PiB-negative). The cut-off was established using Gaussian mixture modeling, as previously described (20).

Statistical Analyses

A chi-square test was used to assess differences in gender across diagnoses. Fisher's exact test was used to compare *APOE* genotype across diagnoses. One-way ANOVAs were used to compare age, education, and OAS score (assessed at the time of olfactory testing) across diagnoses (CN, SCC, MCI, and AD). A one-way ANOVA was also used for the analysis of Trails B time, which was transformed using the Box-Cox procedure (21) to help satisfy the normality assumption of ANOVA. The Kruskal-Wallis test was used to analyze CDR-SOB, MMSE, and BNT scores across diagnoses, while simple logistic regression was used to compare the proportions of correct responses in the OPID-10, OPID-20, OD, and POEM tests across diagnoses. For pairwise comparisons between diagnoses, we used Holm's stepdown method (22) to correct p-values for multiple comparisons. A subgroup analysis of CN participants was run to compare the performance in the POEM test for correctly versus incorrectly identified OPID-10 odors. A logistic mixed effects model was used to estimate the probability of having a "yes" response on the POEM test, while controlling for the effect of correctly identifying each OPID-10 odor (yes versus no) and accounting for the correlation among POEM responses within each participant.

Primary analyses of the proportions of correct responses in the OPID-20, OD, and POEM tests consisted of multiple logistic regression models that controlled for the following effects: diagnosis, age, gender, education level, and interactions between diagnosis and the covariates (age, gender, and education level). We then examined the interaction with the largest p-value (based on a global F test) and determined whether that p-value was less than 0.1. If not, this interaction was removed and the model was re-fit. We continued in this manner until we had a final model that controlled for the effects of diagnosis, age, gender, education level, and interactions whose p-values (based on a global F test) were less than 0.1.

For each proportion of correct responses in the OPID-20, OD, and POEM tests, we then conducted secondary analyses in two parts. In the first part, we fit a multiple logistic regression model for each proportion that controlled for the effects of: diagnosis, MMSE score (to control for overall cognition), the interaction between diagnosis and MMSE score, age, gender, education level, and the significant or marginally significant diagnosis interactions (with age, gender, or education level) that were identified in the primary analysis of that proportion. We then used a global F test to determine whether the interaction between diagnosis and MMSE score was significant at the 0.05 level or marginally significant at the 0.1 level. If not, this interaction was removed and the model was re-fit. Using the same modeling approach, we then fit additional multiple logistic regression models for each proportion that controlled for the effects of each of the following five AD measures, and their interactions with diagnosis, in separate models: BNT score (to control for a naming deficit), OAS score (to control for olfactory awareness), CDR-SOB score (to control for overall function), Trails B time (to control for executive function required to perform the task), and *APOE* genotype (to control for genetic risk), which is a well-established risk factor for Alzheimer's and has been reported to influence odor identification performance. Since no distributional assumptions are required for Trails B time as a covariate, it was not transformed in any of these models

In the second part of our secondary analyses, we fit a multiple logistic regression model for each proportion that controlled for the effects of: diagnosis, the AD measures and their interactions with diagnosis that had significant or marginally significant effects in the first part of our secondary analysis.

For significant main or interaction effects involving diagnosis in both the primary and secondary models for each proportion of correct responses in the OPID-20, OD, and POEM tests, we used Holm's stepdown method to adjust p-values and the Bonferroni correction to adjust 95% confidence intervals (CIs) for multiple comparisons when assessing which pairs of diagnoses significantly differed from one another.

A subgroup analysis of CN participants was run to compare the performance in the POEM test for correctly versus incorrectly identified OPID-10 odors. A logistic mixed effects model was used to estimate the probability of having a "yes" response on the POEM test, while controlling for the effect of correctly identifying each OPID-10 odor (yes versus no) and accounting for the correlation among POEM responses within each participant.

Cognitive trajectory score

For the cognitive trajectory measures, two groups were analyzed:

1. Good POEM Performers (n=88): CN or SCC participants whose POEM Index was *above* the lower bound of the 50% confidence interval of expected POEM Index based on performance on both the OPID-20 and OD tests.
2. Poor POEM Performers (n=24): CN or SCC participants whose POEM Index was *below* the lower bound of the 50% confidence interval of expected POEM Index based on performance on both the OPID-20 and OD tests.

The relationship between change in MMSE, LMII, DSC, and CVLT over time and our odor measures at baseline was modeled using linear mixed effects models with the known modifiers of age, gender, education, and time from baseline visit.

We ran subgroup analyses to compare the performance of good versus poor POEM performers. T-tests were used to compare OAS Sum Score and Trails B time (transformed) at the time of olfactory testing as well as volumetric imaging data and PiB retention between subgroups. Wilcoxon rank-sum tests were used to compare the following test scores at the time of olfactory testing between subgroups: CDR-SOB, MMSE, BNT, CVLT, and UPDRS. The chi-square tests and simple logistic modeling procedures as previously described were used to compare *APOE* genotype between subgroups, as well as the proportions of correct responses in the OPID-10, OPID-20, OD, and POEM tests. In modeling each olfactory measure and cognitive score, all statistical tests were conducted at the 0.05 significance level. In testing hypotheses of pairwise differences between diagnoses for a specific main or interaction effect, we adjusted our significance level of 0.05 for each hypothesis test to correct for multiple comparisons using the correction methods previously described.

All chi-square tests, t-tests, and ANOVA models were implemented in SPSS. Wilcoxon rank-sum and Kruskal-Wallis tests were implemented using `proc npar1way` in SAS Version 9.4. All logistic regression models were implemented in R using the `glm` function in the `stats` package for generalized linear models with the logit link. Linear mixed models were fit using `proc mixed` in SAS Version 9.3.

RESULTS

Demographics of the Cohort

Mean education levels and OAS scores did not significantly differ between the CN, SCC, MCI, and AD participants at the time of olfactory testing, but age significantly differed across these diagnoses (Table 1). As expected, there were highly significant differences between all diagnoses for the CDR-SOB and Trails B tests at the time of olfactory testing (Table 1) as well as the frequency of the *APOE* ϵ 4 allele. MMSE scores significantly differed between each pair of diagnoses, except between MCI and AD diagnoses (Table 1). BNT scores significantly differed between CN and AD participants (Table 1). Participants endorsing specific potential confounding factors during olfactory screening did not perform

significantly differently on the olfactory battery (OD, OPID-10, OPID-20, POEM), the OAS, or NPT when compared to participants who denied these factors (data not shown); hence, all participants were retained for further analysis.

Identification of Odor Percepts

Using simple logistic modeling, we found that OPID-20 accuracy significantly differed between each pair of diagnoses (Table 2, Figure 2A). In addition, OPID-10 accuracy significantly differed between AD participants and those in the CN and SCC groups, and also between MCI participants and those in the CN and SCC groups (Table 2, Figure 2A).

In multiple regression models, and paralleling previous studies of a smell identification outcome and diagnosis (9, 23), we found that controlling for the effects of age, education, and gender, there were significant interactions between diagnosis and age (global p (p using global F test) < 0.0001), between diagnosis and gender (global $p = 0.03$), and between diagnosis and education level (global $p = 0.04$) on the proportion of correct trials for the OPID-20 test. The estimated log odds ratio (LOR) of correct identification for the SCC versus the MCI diagnoses for males was 0.658 less than that for females (adjusted p (p using Holm's stepdown method) $= 0.016$, 95% CI: $(-1.236, -0.081)$). For every additional year of participant age, (1) the estimated LOR of correct identification for the SCC diagnosis compared with the AD diagnosis significantly decreases by 0.116 (adjusted $p < 0.0001$, 95% CI: $(-0.179, -0.053)$), (2) the estimated LOR of correct identification for the MCI diagnosis compared with the AD diagnosis significantly decreases by 0.122 (adjusted $p < 0.0001$, 95% CI: $(-0.189, -0.055)$), and (3) the estimated LOR of correct identification for the CN diagnosis compared with the AD diagnosis significantly decreases by 0.097 (adjusted $p < 0.0001$, 95% CI: $(-0.160, -0.034)$). For every additional year of education, (1) the estimated LOR of correct identification for the MCI diagnosis compared with the CN diagnosis marginally significantly decreases by 0.117 (adjusted $p = 0.068$, 95% CI: $(-0.24, 0.005)$).

In the first part of our secondary analysis, we found that controlling for the effects of diagnosis, age, gender, and education, the main effect of OAS test score on the proportion of correctly identified trials in the OPID-20 test was significant (global $p < 0.0001$), along with the interactions between diagnosis and MMSE score (global $p < 0.0001$) and between diagnosis and Trails B time (global $p < 0.0001$).

In the second part of our secondary analysis, the main effect of OAS test score on the proportion of correctly identified trials in the OPID-20 test remained significant (global $p < 0.001$), controlling for the effects of diagnosis, age, gender, education, MMSE score, and Trails B time. Due to significant interactions between diagnosis and MMSE score (global $p < 0.0001$), as well as diagnosis and Trails B time (global $p < 0.0001$), the effects of MMSE and Trails B performance on the proportion of correctly identified trials in the OPID-20 test are significantly modified by diagnosis.

Percepts of Odor Episodic Memory (POEM)

Simple logistic regression modeling revealed significant differences in POEM test performance between AD participants and those in the CN and SCC groups (Table 2, Fig. 2B). Significant main effects of diagnosis (global $p = 0.003$) and age (global $p = 0.02$) were

revealed when multiple logistic regression was used to model the proportion of correct trials in the POEM test, controlling for the effects of gender and education. Relative to AD participants, (1) the log odds of correctly distinguishing previously presented odors were significantly higher by 0.556 (adjusted $p = 0.002$; 95% CI: (0.146,0.965)) for the CN group and by 0.435 (adjusted $p = 0.022$; 95% CI: (0.032,0.838)) for the SCC group. For a given diagnosis, gender and education level, the log odds of correct episodic recall decreased by 0.01 (global $p = 0.02$; 95% CI: (-0.019, -0.002)) for every additional year of participant age.

In the first part of our secondary analysis, none of the six AD measures, or their interactions with diagnosis, had a significant or marginally significant effect on the proportion of correct trials in the POEM test, controlling for the effects of diagnosis, age, gender, and education. Therefore, we did not proceed to the second part of our secondary analysis. In addition, to determine whether correctly naming the odor in the OPID-10 affected the degree of encoding, i.e., no feedback was provided, we used a logistic mixed effects model to compare the proportion of correctly remembered odors in the POEM test stratified by whether they were correctly or incorrectly identified in the OPID-10. In 18 CN subjects, where the OPID-10 proportion correct was 0.5 or 0.6 to balance the number of correct and incorrect trials, no difference was noted ($p = 0.45$).

Odor Discrimination

Simple logistic regression modeling of the OD outcome showed significant differences in performance between the CN and AD diagnoses (Table 2, Fig. 2C). Modeling the proportion of trials with correct discrimination revealed significant main effects of diagnosis (global $p=0.01$) and age (global $p = 0.0004$) controlling for covariate effects. Relative to AD participants, the log odds of correct discrimination was significantly higher for the MCI group by 0.608 (adjusted $p = 0.043$, 95% CI: (0.00, 1.217)) and by 0.666 for the CN group (adjusted $p = 0.011$, 95% CI: (0.1, 1.232)), controlling for all other covariate effects. For a given diagnosis, gender and education level, the log odds of correct discrimination significantly decreased by 0.023 (global $p = 0.0004$, 95% CI: (-0.036,-0.011)) for every additional year of participant age.

In the first part of our secondary analysis, we found that controlling for the effects of diagnosis, age, gender, and education, the main effect of Trails B time on the proportion of trials with correct discrimination was significant (global $p < 0.0001$). In the second part of our secondary analysis, we found that the log odds of correct discrimination significantly decreased by 0.006 (global $p = 0.0002$, 95% CI: (-0.009,-0.003)) with each unit increase in Trails B score, controlling for the effects of diagnosis, age, gender, and education.

Hippocampal Volume and Entorhinal Thickness, but not Amyloid Deposition, Correlate with OPID Performance

To adjust for the variable timing of high resolution MRI imaging relative to olfactory testing, in the subgroup of participants who underwent MRI imaging ($n=49$; 11 CN; 27 SCC; 9 MCI; 2 AD), multiple linear regression models were fit with an interaction term for imaging timing x measurement (e.g. entorhinal thickness, adjusted bilateral hippocampal volume) in

addition to imaging timing, and imaging measurement variables. None of the models were significant for imaging timing or the imaging timing/ measurement interaction term. When we control for the main effect of imaging timing, we find significant partial correlations between OPID-20 score and adjusted bilateral hippocampal volume ($r = 0.534$, $p < 0.0001$) (Fig. 3A) and entorhinal thickness ($r = 0.321$, $p = 0.026$), consistent with previous findings using the UPSIT (24). Significant partial correlations, adjusted for the main effect of imaging timing, were also observed between OPID-10 score and adjusted bilateral hippocampal volume ($r = 0.323$, $p = 0.026$) and entorhinal thickness ($r = 0.313$, $p = 0.03$). Performance on the POEM or OD tests did not significantly correlate with either volumetric measure in an unadjusted or adjusted analysis.

In the subgroup of participants who underwent amyloid PET imaging ($n=41$; 9 CN; 22 SCC; 8 MCI; 2 AD), we found no significant correlations between FLR score and performance on the OD, OPID-20, or POEM tests across or within diagnoses. To adjust for the variable timing of amyloid-imaging relative to olfactory testing, in the subgroup of participants who underwent amyloid PET imaging, a multiple linear regression model was fit with an interaction term for imaging timing x measurement (e.g. FLR score) in addition to imaging timing, and FLR score variables. None of the models were significant for imaging timing or the imaging timing/ measurement interaction term. When we control for the main effect of the timing of amyloid imaging, we do not observe any statistically significant partial correlations between FLR score and any of the olfactory measures.

Identification of Poor POEM Performers (PPPs) among CN and SCC Participants

The episodic recognition of odors as previously (or not previously) presented in the OPID-20 is a complex task that relies on both the ability to discriminate and identify odor percepts. Consistent with this, performance on the POEM and the OD and OPID-20 tests were substantially correlated across our sample. We hypothesized that individuals who had pathophysiological changes in the olfactory/entorhinal cortical neural systems would perform less well on the episodic memory component relative to the identification and discrimination tasks, even early on in disease. In order to identify individuals with specific deficits in *episodic memory* of odor percepts, we developed two prediction models, one for POEM index as a function of the OPID-20, and another for POEM index as a function of OD. CN and SCC participants with POEM indices less than the lower bounds of the 50% confidence interval of the predicted scores from both of these models were identified as poor POEM performers (PPPs, $n=24$; 21%) and the remainder were identified as good POEM performers ($n=88$). This 50% threshold was based on our initial observations that POEM indices of the lowest quartile of CN participants approximated the range of values of the AD participants (whose performance was no better than chance).

Good performers and poor performers on the POEM test did not significantly differ from one another on their performance of the discrimination (78.2% correct vs. 77.4%, $p=0.77$) nor on other demographic variables, with the exception of the CVLT test (73.4% correct vs. 66.4%, $p=0.01$). However, PPPs did perform significantly worse on the identification tasks (OPID-10: 70% correct vs. 58.8%, $p=.001$; OPID-20: 76.3% correct vs. 69.4%, $p=0.002$).

Enrichment of *APOE* $\epsilon 4$ and Reduced Entorhinal Thickness in PPPs

We next examined biomarkers related to underlying AD related pathobiology in clinically normal individuals who were good versus poor POEM performers. Using a Pearson chi-square analysis, PPPs were significantly more likely to carry the *APOE* $\epsilon 4$ allele than good POEM performers ($p=0.011$; $\phi = 0.25$; Figure 4A). Entorhinal thickness in PPPs was marginally significantly decreased ($p=0.098$; $d = 0.56$; Figure 4B) relative to the good POEM performers. The time window between MRI scans and olfactory testing was not significantly different in the good and poor POEM performers ($p=0.208$). Mean adjusted bilateral hippocampal volume and PiB retention did not significantly differ between good POEM performers and PPPs, and PiB retention did not significantly correlate with the performance on the olfactory tests themselves. Moreover, no significant differences in FLR score were found between good and PPPs when it was considered as a continuous ($p=0.423$) or a binary outcome thresholded at $FLR=1.2$ ($p=0.131$). No significant difference was found in the length of the time between PiB imaging and olfactory testing in the good vs. poor POEM performers ($p=0.188$).

Worse Trajectory of Episodic Memory in Clinically Normal Individuals with Selective Odor Memory Loss

Based on our hypothesis of the predictive value of a selective episodic odor memory deficit (rather than a more global deficit in olfactory function), trajectories of MMSE, DSC, LM II, and CVLT and a composite score equally weighting all 4 tests (24) were modeled for all clinically normal participants (diagnosis = CN or SCC), who had 3 or more visits with available NPT ($n=111$). Participants with a selective odor memory deficit (PPP, $n=24$) were compared with the remainder of this sample (good POEM performers, $n=88$). For CN and SCC participants with good POEM performance (i.e., commensurate with their performance on other odor tasks, as described above), all scores were modeled, since it could be inferred that these participants also had good POEM performance at their earlier visit dates. In additional analyses, we truncated the scores at the time of testing to assess neuropsychological test trajectories in a known state of good POEM performance. For PPPs, trajectories of cognitive scores were modeled from the time of olfactory testing and forward, since the POEM status of these participants at the earlier time points could not be inferred.

Consistent with the previously reported practice effect (25), the trajectory of LM II scores over time had a significant positive slope for the good POEM performing group ($n=88$, $\beta=0.48$, $p<0.0001$) (Table 3). By contrast, the trajectory of LM II score over time for the PPPs did not differ significantly from zero ($n=24$, $\beta=-0.20$, $p=0.516$), suggesting practice effect may have been absent in the group of subjects who showed selective impairments on the episodic olfactory memory test. Consistent with our hypothesis, the slopes of LMII scores over time for good POEM performers were significantly better ($p=0.032$) than poor POEM performers. Use of all longitudinal data for good POEM performers did not change any of these results. The trajectory of LM II scores for the CN/SCC participants in the lowest quartile of POEM index scores was significantly different ($p=0.010$) from the upper three quartiles of CN/SCC participants ($n=88$, $\beta= 0.50$, $p<0.0001$ vs. $n=24$, $\beta=-0.28$, $p=0.368$). By contrast, OAS ($p=0.577$), BNT ($p=0.62$), Trails-B (transformed) ($p=0.92$), and trajectories of MMSE ($p=0.73$), DSC ($p= 0.78$), and CVLT ($p= 0.25$) did not significantly

differ between good and poor POEM performers. Analysis of prospective neuropsychological data from the time of olfactory testing did not show significant differences between the two groups (mean follow-up is 1.5 years). Taken together, these data indicate that *selective* poor performance on the POEM test (relative to the other olfactory measures) in clinically normal individuals is associated with a worse trajectory of LM II scores.

DISCUSSION

In this study, we demonstrate significant differences in OPID-20 performance between all diagnostic groups representing the spectrum of AD. Moreover, we find worse OPID performance is associated with reduced adjusted hippocampal volume and thinner entorhinal cortices, paralleling previous reports of significant relationships between the UPSIT and adjusted hippocampal volume and entorhinal volume (23, 26). We observe these associations in spite of the fact that the MRI imaging was performed an average of nearly two years prior to the olfactory testing. The lack of association between OPID and amyloid imaging may be due to the earlier date, on average, of the amyloid scans relative to olfactory testing. In secondary analyses, we find that the performance on the OPID-20 has multiple interactions with age, education, global cognition, executive functioning and the subject's subjective awareness of odors as well as gender, the latter was previously reported for the UPSIT (27). No association was found between OPID-20 and *APOE* genotype. The OD measure discriminated less between the diagnostic categories — significantly differing only between individuals who were cognitively normal and who had AD dementia. The OD measure was also influenced by age, gender, education, and executive function. These multiple associations potentially complicate the interpretation of odor identification and odor discrimination deficits as a biomarker in clinically normal populations. Performance on our novel POEM index of odor memory is significantly different between both CN and SCC participants and patients with dementia. In contrast to the OPID and the OD, our secondary analyses reveal that the POEM index is only associated with age, the principal risk factor for sporadic AD — although these interaction tests may be underpowered. The AD patients perform at chance on this test of odor memory; the floor effect of the POEM tests suggests that it will not be useful to follow disease progression in the dementia stage of AD, but offers good dynamic range in the preclinical population.

While the lack of association of POEM with the other demographic and neuropsychological measures is promising as an independent AD biomarker, smell function measurements in normal participants vary over several orders of magnitude (28, 29). The mechanism(s) which influence this broad performance range likely relate to individual variations in genetics and physiology. We hypothesized that normalizing for each participant's intrinsic olfactory function (based on their performance on the identification and discrimination tasks) would yield an olfactory-based episodic memory score with greater specificity for dysfunction of neural systems commonly impaired in early AD, namely the entorhinal cortex and dedicated olfactory processing regions (8). Indeed, clinically normal participants with observed odor *memory* scores below the scores predicted by both their odor *identification* and *discrimination* scores were enriched in three established AD biomarkers: *APOE* ϵ 4 alleles, entorhinal cortical thinning, and worse trajectory on an episodic memory task. By contrast,

clinically normal participants in the bottom quartile of OPID-20 performance were not enriched in any of these AD biomarkers, and the bottom quartile of POEM only had a worse cognitive trajectory. Both the presence of *APOE* ϵ 4 allele and thinner entorhinal cortices are well accepted markers of risk for developing AD (30).

Limitations of our study include significant differences in age and gender across the diagnostic categories; we adjusted for these potential confounders in our multivariate logistic regression models. In addition, only a subset of our cohort had available biomarkers for amyloid deposition, entorhinal thickness, hippocampal volume, and *APOE* genotype. Many studies have found diminished smell identification performance in patients with Parkinson's disease (PD) and in those at elevated risk for developing PD (11). In addition, patients with Lewy Body Dementia (LBD) also exhibit significant odor identification deficits (31). Future studies will be directed at determining the specificity of PPP in patients with symptomatic PD and LBD.

Following confirmation in longer prospective longitudinal analyses and in other cohorts, selective odor episodic odor memory loss, i.e. PPP status, may identify a subset of clinically normal participants at greater risk for developing the progressive memory symptoms of AD. Combining olfactory assessments (identification and POEM) with genetic, imaging, and molecular biomarkers for AD pathology may yield an enhanced AD risk profile (32) to identify clinically normal elderly participants, with an increased probability of cognitive decline in the next two years, for clinical trials. In addition, olfactory assessments are inexpensive and well-tolerated, so they might play an important role in cost-effective screening for preclinical AD in the general population.

Acknowledgments

We thank the participants and their families for their time and efforts. We thank Jeanette Gunther for coordinating the ADRC longitudinal cohort, Kyleen Swords, Stephanie Camhi, Amy Zoller, Frances Hatling, Larissa Collins, and Sandra Yan for testing participants. We thank Matthew Growdon, M.D. and Gad Marshal, M.D. for providing constructive comments. Supported by NIH DP2-OD000662 (awarded to M.W.A), P30-AG036449 (awarded to R.A.S.), P50-AG005134 (awarded to B.T.H.), T32NS048005 (awarded to R.A.B.), the Wilkens Foundation (awarded to M.W.A), MGH ECOR grant (awarded to M.W.A), and the Harvard NeuroDiscovery Center.

Bibliography

1. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of 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(3):280–92. [PubMed: 21514248]
2. Carrillo MC, Blackwell A, Hampel H, Lindborg J, Sperling R, Schenk D, et al. Early risk assessment for Alzheimer's disease. *Alzheimers Dement.* 2009; 5(2):182–96. [PubMed: 19328456]
3. Dickerson BC, Stoub TR, Shah RC, Sperling RA, Killiany RJ, Albert MS, et al. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology.* 2011; 76(16):1395–402. [PubMed: 21490323]
4. Hedden T, Mormino EC, Amariglio RE, Younger AP, Schultz AP, Becker JA, et al. Cognitive profile of amyloid burden and white matter hyperintensities in cognitively normal older adults. *J Neurosci.* 2012; 32(46):16233–42. [PubMed: 23152607]
5. Toledo JB, Xie SX, Trojanowski JQ, Shaw LM. Longitudinal change in CSF Tau and A β biomarkers for up to 48 months in ADNI. *Acta Neuropathol.* 2013; 126(5):659–70. [PubMed: 23812320]

6. Amariglio RE, Frishe K, Olson LE, Wadsworth LP, Lorus N, Sperling RA, et al. Validation of the Face Name Associative Memory Exam in cognitively normal older individuals. *J Clin Exp Neuropsychol.* 2012; 34(6):580–7. [PubMed: 22765048]
7. Hyman BT, Arriagada PV, Van Hoesen GW. Pathologic changes in the olfactory system in aging and Alzheimer's disease. *Ann N Y Acad Sci.* 1991; 640:14–9. [PubMed: 1776730]
8. Kovacs T, Cairns NJ, Lantos PL. Olfactory centres in Alzheimer's disease: olfactory bulb is involved in early Braak's stages. *Neuroreport.* 2001; 12(2):285–8. [PubMed: 11209936]
9. Devanand DP, Lee S, Manly J, Andrews H, Schupf N, Doty RL, et al. Olfactory deficits predict cognitive decline and Alzheimer dementia in an urban community. *Neurology.* 2015; 84(2):182–9. [PubMed: 25471394]
10. Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiol Behav.* 1984; 32(3):489–502. [PubMed: 6463130]
11. Albers MW, Tabert MH, Devanand DP. Olfactory dysfunction as a predictor of neurodegenerative disease. *Curr Neurol Neurosci Rep.* 2006; 6(5):379–86. [PubMed: 16928347]
12. Doty RL, Shaman P, Applebaum SL, Giberson R, Siksorski L, Rosenberg L. Smell identification ability: changes with age. *Science.* 1984; 226(4681):1441–3. [PubMed: 6505700]
13. Okereke OI, Pantoja-Galicia N, Copeland M, Hyman BT, Wanggaard T, Albert MS, et al. The SIST-M: predictive validity of a brief structured clinical dementia rating interview. *Alzheimer Dis Assoc Disord.* 2012; 26(3):225–31. [PubMed: 21986342]
14. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res.* 1982; 17(1):37–49. [PubMed: 7183759]
15. Rosen WG, Terry RD, Fuld PA, Katzman R, Peck A. Pathological verification of ischemic score in differentiation of dementias. *Ann Neurol.* 1980; 7(5):486–8. [PubMed: 7396427]
16. Tabert MH, Liu X, Doty RL, Serby M, Zamora D, Pelton GH, et al. A 10-item smell identification scale related to risk for Alzheimer's disease. *Ann Neurol.* 2005; 58(1):155–60. [PubMed: 15984022]
17. Smeets MA, Schifferstein HN, Boelema SR, Lensvelt-Mulders G. The Odor Awareness Scale: a new scale for measuring positive and negative odor awareness. *Chem Senses.* 2008; 33(8):725–34. [PubMed: 18622009]
18. Weintraub S, Salmon D, Mercaldo N, Ferris S, Graff-Radford NR, Chui H, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychological test battery. *Alzheimer Dis Assoc Disord.* 2009; 23(2):91–101. [PubMed: 19474567]
19. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry.* 1982; 140:566–72. [PubMed: 7104545]
20. Sperling RA, Laviolette PS, O'Keefe K, O'Brien J, Rentz DM, Pihlajamaki M, et al. Amyloid deposition is associated with impaired default network function in older persons without dementia. *Neuron.* 2009; 63(2):178–88. [PubMed: 19640477]
21. Box GE, Cox DR. An Analysis of Transformations. *Journal of the Royal Statistics Society, Series B.* 1964; 26:211–34.
22. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics.* 1979; 6:65–70.
23. Devanand DP, Tabert MH, Cuasay K, Manly JJ, Schupf N, Brickman AM, et al. Olfactory identification deficits and MCI in a multi-ethnic elderly community sample. *Neurobiol Aging.* 2008
24. Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol.* 2014; 71(8):961–70. [PubMed: 24886908]
25. Mormino EC, Betensky RA, Hedden T, Schultz AP, Amariglio RE, Rentz DM, et al. Synergistic effect of beta-amyloid and neurodegeneration on cognitive decline in clinically normal individuals. *JAMA Neurol.* 2014; 71(11):1379–85. [PubMed: 25222039]

26. Growdon ME, Schultz AP, Dagley AS, Amariglio RE, Hedden T, Rentz DM, et al. Odor identification and Alzheimer disease biomarkers in clinically normal elderly. *Neurology*. 2015; 84(21):2153–60. [PubMed: 25934852]
27. Doty RL, Applebaum S, Zusho H, Settle RG. Sex differences in odor identification ability: a cross-cultural analysis. *Neuropsychologia*. 1985; 23(5):667–72. [PubMed: 4058710]
28. Cain WS, Gent JF. Olfactory sensitivity: reliability, generality, and association with aging. *J Exp Psychol Hum Percept Perform*. 1991; 17(2):382–91. [PubMed: 1830082]
29. Hasin-Brumshtein Y, Lancet D, Olender T. Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet*. 2009; 25(4):178–84. [PubMed: 19303166]
30. Jagust W. Vulnerable neural systems and the borderland of brain aging and neurodegeneration. *Neuron*. 2013; 77(2):219–34. [PubMed: 23352159]
31. Olichney JM, Murphy C, Hofstetter CR, Foster K, Hansen LA, Thal LJ, et al. Anosmia is very common in the Lewy body variant of Alzheimer’s disease. *J Neurol Neurosurg Psychiatry*. 2005; 76(10):1342–7. [PubMed: 16170073]
32. Devanand DP, Liu X, Tabert MH, Pradhaban G, Cuasay K, Bell K, et al. Combining early markers strongly predicts conversion from mild cognitive impairment to Alzheimer’s disease. *Biol Psychiatry*. 2008; 64(10):871–9. [PubMed: 18723162]

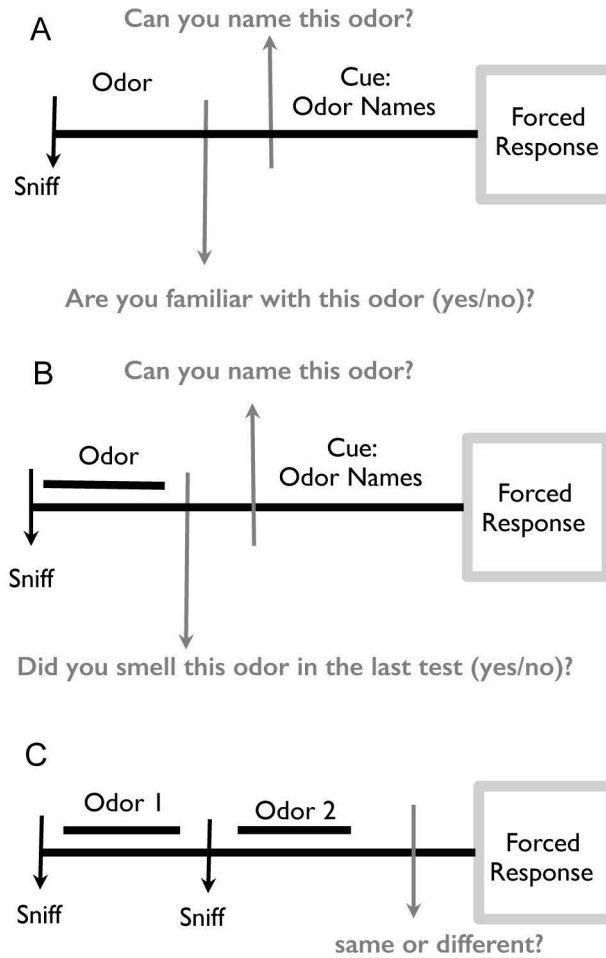


Figure 1. Schematics of Olfactory Tests. (A) 10-item Qdor Percept Identification (OPID-10) test: The participant was cued on each trial by the software to prepare to sniff bilaterally and sample the 2 s delivery of an odor. Following odor presentation, participants were asked if they were *familiar* with the odor (Yes/No). During the instruction phase of the test, the research assistant explicitly indicated that familiarity with the odor percept did not require a semantic label. Following the odor familiarity response, participants were presented with 4 odor names and asked to choose which odor name associated with their memory of the odor percept experienced at the start of the trial. (B) 20-item percept of odor episodic memory (POEM) and identification (OPID-20) tests were administered after a 10 minute delay. Odor presentation is as described for the OPID-10. Immediately following odor presentation, participants were asked if the odor had been presented in the earlier odor percept identification test (Yes/No). During the instruction phase of the test, the research assistant explicitly indicated that New/Old designation referred to the testing session and not to a broader lifetime exposure. Following this measure of episodic odor percept memory, the naming is as described in A. (C) The 12-item odor discrimination (OD) test. The participant was cued on each trial to prepare to sniff bilaterally and sample the delivery of two odors presented consecutively for 2 seconds each. Participants were then asked if the two odors

presented were the same or different. Half of the trials were the “Same”, and the odors included were a predetermined selection of the odors listed above.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

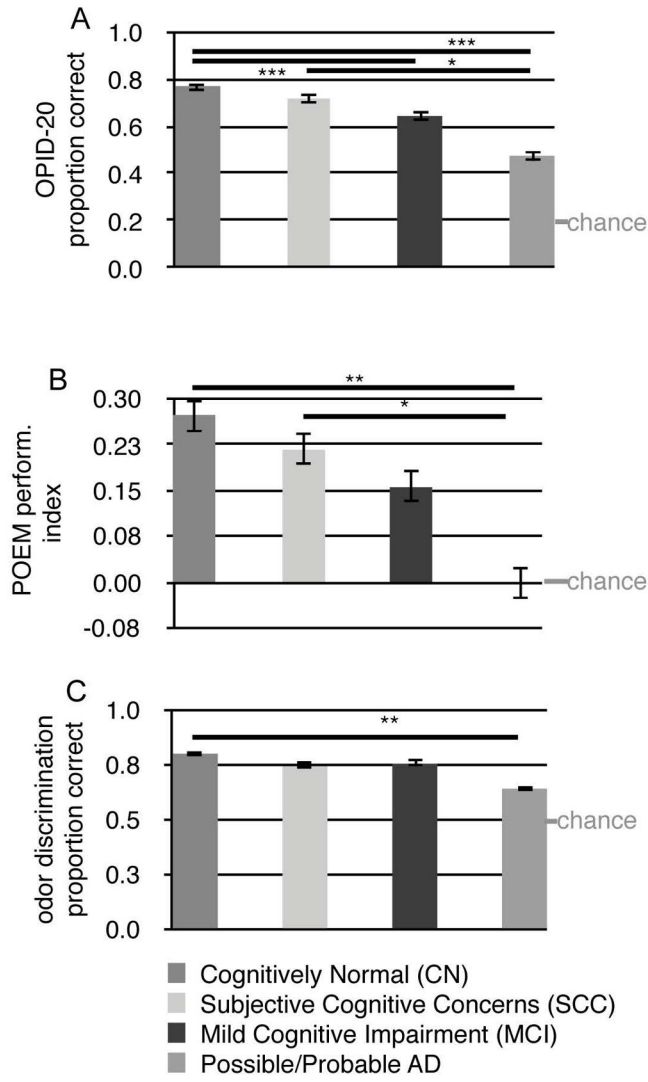


Figure 2. Odor Percept Identification, POEM index, and Odor Discrimination Across Disease Categories. (A) Proportion of the 20-item odor percepts identified correctly and S.E.Ms for each diagnosis. (B) POEM Performance Index ((# correct “New” or “Old” responses) - (# of incorrect responses) / 20 total responses) for each diagnosis. A score of 0 corresponds to participants performing at chance. (C) Proportion of odors correctly discriminated. Means and S.E.M. are shown. Bars indicate significant differences between diagnoses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$.

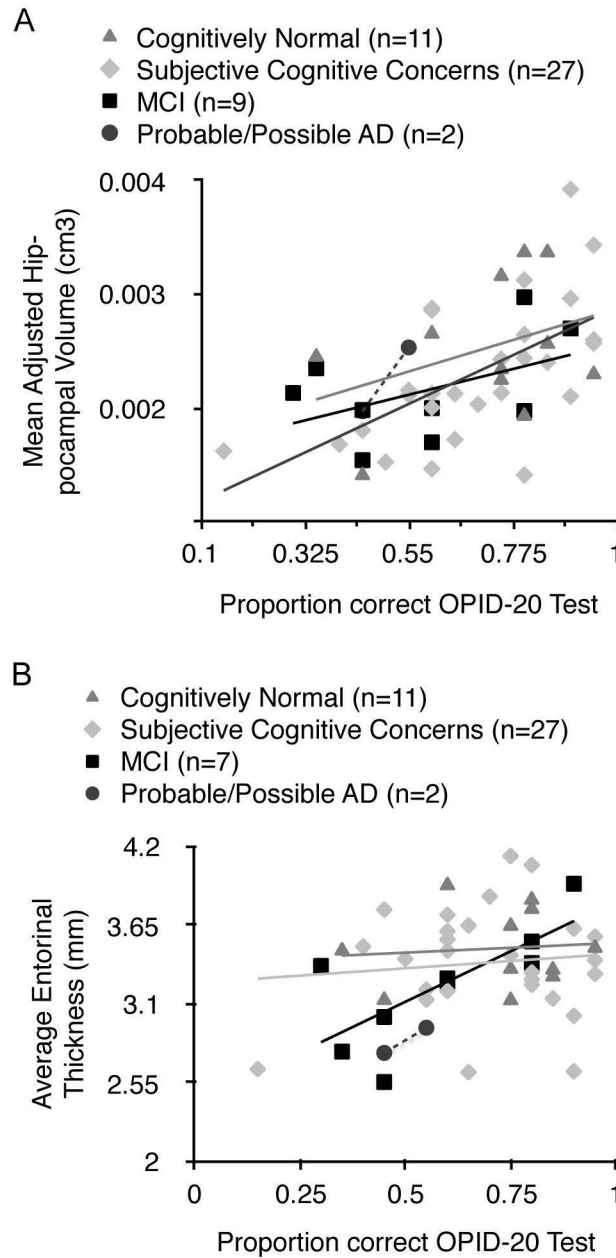


Figure 3.

Adjusted hippocampal volume and entorhinal cortical thickness correlates with OPID-20 score. (A) Bilateral hippocampal volumes determined from MRI were averaged and adjusted for head size. Scores on the OPID-20 test are plotted against the adjusted hippocampal volume, ($r = 0.534$.; $p < 0.0001$) (B) Bilateral entorhinal thickness determined from MRI were averaged. Scores on the OPID-20 test are plotted against entorhinal cortical thickness. ($r = 0.321$.; $p = 0.026$).

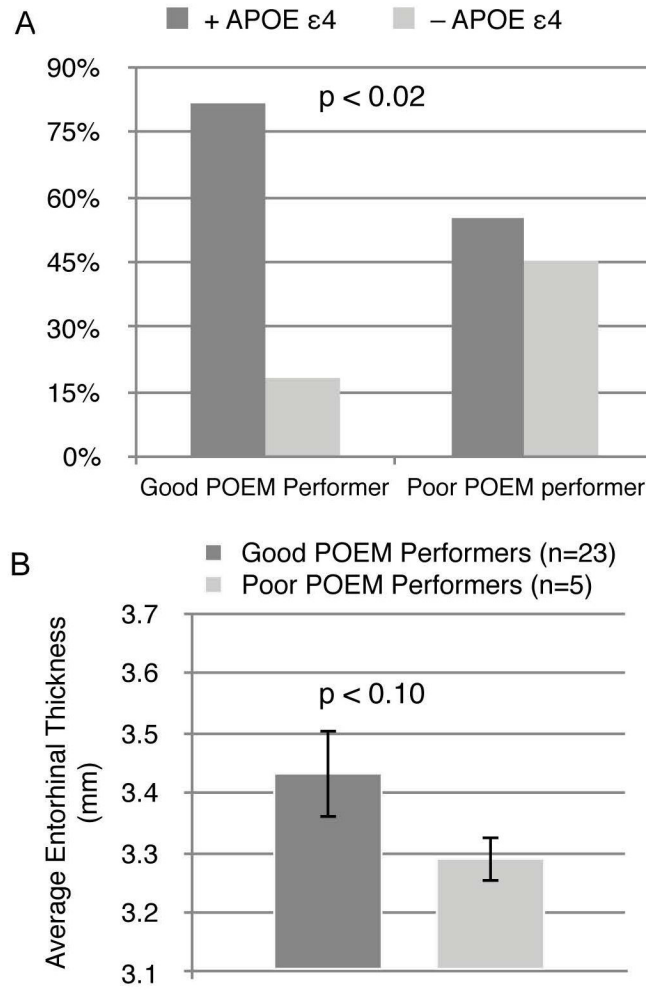


Figure 4. Poor performance on the POEM normalized to the OPID-20 and OD outcomes identifies CN/SCC participants enriched with AD biomarkers. (A) The CN/SCC participants who were good POEM performers (n = 87) had significantly less frequency of the *ApoE* ε4 allele (18%) relative the PPPs (45%; n = 20; p = 0.011) (B) In the subset of olfactory tested CN/SCC participants who underwent volumetric MRI, average entorhinal thickness of good POEM performers (n = 23) was greater than the poor POEM performers (n = 5) (p = 0.098).

Table 1

Participant Demographics

| Demographic Variable | Cognitively Normal (CN) (n=70) | Subjective Cognitive Concerns (SCC)(n=74) | Mild Cognitive Impairment (MCI) (n=29) | Probable/Possible AD (n=10) | p-value |
|---------------------------------------|--------------------------------|---|--|-----------------------------|---------|
| Mean Age (SE) ^d | 75 (1.01) | 77 (0.93) | 80 (1.52) | 77 (2.60) | 0.045 |
| Mean Education (SE) | 17 (0.26) | 16 (0.29) | 16 (0.48) | 15 (1.05) | 0.47 |
| Sex (% female) | 71% | 59% | 48% | 40% | 0.07 |
| ApoE4 Status (%+4) [*] | 24% | 23% | 40% | 100% | 0.0002 |
| Mean CDR-SOB (SE) ^{***, b} | 0.01 (0.01) | 1.02 (0.08) | 1.52 (0.18) | 4.50 (0.19) | <0.0001 |
| Mean MMSE Score (SE) ^{***c} | 29.45 (0.10) | 28.95 (0.16) | 27.62 (0.28) | 24.71 (1.32) | <0.0001 |
| Mean BNT Score (SE) ^{***d} | 28.66 (0.20) | 27.99 (0.25) | 26.76 (0.74) | 24.43 (1.66) | 0.0036 |
| Mean Trails B Time (SE) ^{†e} | 62.43s (3.2s) | 75.66s (3.95s) | 109.03s (9.04s) | 167.00s (29.14s) | <0.0001 |
| Mean OAS Sum Score (SE) [‡] | 69 (1.04) | 71 (1.21) | 71 (1.82) | 70 (4.70) | 0.53 |

^aOverall p value from one-way ANOVA. Significant pairwise difference between CN and MCI (p<0.05).

^bOverall p value from Kruskal-Wallis test. All pairwise differences between diagnoses are significant (p<0.02)

^cOverall p value from Kruskal-Wallis test. Significant pairwise differences between CN and SCC (p<0.03), CN and MCI (p<0.0001), CN and AD (p<0.02), SCC and MCI (p<0.001), SCC and AD (p<0.001), and SCC and AD (p<0.01)

^dOverall p value from Kruskal-Wallis test. Significant pairwise difference between CN and AD (p<0.02)

^eOverall p value from one-way ANOVA on transformed scale. Significant pairwise differences (on transformed scale) between CN and AD (p<0.0001), SCC and MCI (p<0.001), SCC and AD (p<0.001), CN and SCC (p<0.02), and CN and MCI (p<0.0001)

^{*}For this measure, n=67 for CN, n=66 for SCC, n=25 for MCI, and n=7 for Probable/Possible AD.

^{**}For this measure, n=70 for CN, n=73 for SCC, n=29 for MCI, and n=9 for Probable/Possible AD.

^{***}For this measure, n=67 for CN, n=73 for SCC, n=28 for MCI, and n=7 for Probable/Possible AD.

[†]For this measure, n=67 for CN, n=73 for SCC, n=28 for MCI, and n=6 for Probable/Possible AD.

[‡]For this measure, n=70 for CN, n=73 for SCC, n=29 for MCI, and n=10 for Probable/Possible AD.

Abbreviations: Cognitively Normal (CN), Subjective Cognitive Concerns (SCC), MCI = mild cognitive impairment, AD = Alzheimer's disease, SE = standard error of the mean, CDR-SOB = clinical dementia rating-sum of boxes, MMSE = mini mental status examination, BNT = Boston Naming Test, OAS = Odor Awareness Scale.

Olfactory Test Scores

Table 2

| Olfactory Outcome | Cognitively Normal (CN) (n=70) | Subjective Cognitive Concerns (SCC) (n=74) | Mild Cognitive Impairment (MCI) (n=29) | Probable/Pos sible AD (n=10) | p-value |
|---|-----------------------------------|---|---|------------------------------|--------------------|
| Mean OPID-10 Prop. Correct (SE) ^a | 0.69 (0.02) | 0.67 (0.02) | 0.56 (0.03) | 0.52 (0.05) | < 0.0001 |
| Mean OPID-20 Prop. Correct (SE) ^b | 0.76 (0.01) | 0.72 (0.01) | 0.64 (0.02) | 0.47 (0.04) | < 0.0001 |
| Mean POEM Prop. Correct (SE) ^c | 0.64 (0.01) | 0.60 (0.01) | 0.58 (0.02) | 0.5 (0.04) | 0.001 |
| Mean POEM Index Prop. Correct (SE) ^d | 0.27 (0.03) | 0.21 (0.03) | 0.16 (0.04) | 0.00 (0.07) | 0.001 ^e |
| Mean OD Prop. Correct (SE) ^f | 0.80 (0.01) | 0.75 (0.01) | 0.76 (0.02) | 0.64 (0.04) | 0.008 |

^aOverall p value from simple logistic model. Significant pairwise differences between CN vs MCI (log odds ratio [LOR] = 0.56; p=0.0005), CN vs Probable AD (LOR = 0.71; p=0.004), MCI vs SCC (LOR = -0.48; p=0.004), and SCC vs Probable AD (LOR = 0.62; p=0.01).

^bOverall p value from simple logistic model. Significant pairwise differences between CN vs MCI (LOR = 0.60; p<0.0001), CN vs SCC (LOR = 0.24; p=0.005), CN vs Probable AD (LOR = 1.30; p<0.0001), MCI vs SCC (LOR = -0.35; p=0.00), MCI vs Probable AD (LOR = 0.70; p<0.0001), and SCC vs Probable AD (LOR = 1.06; p<0.0001).

^cOverall p value from simple logistic model. Significant pairwise differences between CN vs Probable AD (LOR = 0.56; p=0.001) and SCC vs Probable AD (LOR = 0.42; p=0.03).

^dDifference between estimated proportion of correct responses and estimated proportion of incorrect responses in POEM test

^eBased on significance test in logistic model with proportion of correct responses in POEM test as response variable

^fOverall p value from simple logistic model. Significant pairwise differences between CN vs Probable AD (LOR = 0.80; p=0.001) and CN vs SCC (LOR = 0.30; p=0.049).

Abbreviations: Prop. = Proportion, SE= Standard Error of the Mean, Cognitively Normal (CN), Subjective Cognitive Concerns (SCC), MCI = mild cognitive impairment, AD = Alzheimer’s disease.

Table 3

Cognitive Trajectories of Good vs. Poor POEM Performers

| Cognitive test | Good POEM Performers | | | Poor POEM Performers | | |
|----------------|----------------------|------|---------|----------------------|------|---------|
| | slope | S.E. | p-value | slope | S.E. | p-value |
| MMSE | | | | | | |
| model 1 | 0.04 | 0.02 | 0.092 | | | |
| model 2 | | | | -0.17 | 0.14 | 0.239 |
| LM2 | | | | | | |
| model 1 | 0.48 | 0.07 | <0.0001 | | | |
| model 2 | | | | -0.20 | 0.31 | 0.516 |

LM II z scores calculated for CN/SCC participants that underwent olfactory testing once and had over 3 years of annual neuropsychological testing. Model 1 uses NP scores up until time of olfactory testing in the good POEM performers and deletes remaining data. Model 2 uses NP scores beginning at time of olfactory testing in poor POEM performers (PPP). Abbreviations: MMSE = mini mental status examination. LM2 = delayed recall of logical memory test.