A Normative Modeling Based Analysis of Amyloid, Cognition, and Tau in Preclinical Alzheimer's Disease

Zirui Tao[1] Ronak R. Mehta [1] Seong Jae Hwang [1] Rebecca Koscik [1 2 3] Erin Jonaitis [1 2 3] Sterling C. Johnson [1 2 3] Vikas Singh [1 2]

[1]University of Wisconsin - Madison [2]Wisconsin Alzheimer's Disease Research Center [3]Wisconsin Alzheimer's Institute

Wisconsin Registry for Alzheimer's Prevention
UNIVERSITY OF WISCONSIN
SCHOOL OF MEDICINE AND PUBLIC HEALTH

Wisconsin Alzheimer's Institute
UNIVERSITY OF WISCONSIN
SCHOOL OF MEDICINE AND PUBLIC HEALTH

INTRODUCTION

Problem.

Given cognitive scores and amyloid measurement, we seek to assess the extent to which their statistical association is correlated with tau measurements in the preclinical AD stages

Importance.

Cognitive scores, amyloid and tau measurements are important for understanding pre-clinical AD. But the temporal trends of these biomarkers are different. What is the extent to which temporal statistical patterns between cognitive tests and amyloid load help predict an individual's tau positivity?

A potential challenge.

The cohort is highly heterogeneous and the statistical signal is difficult to detect

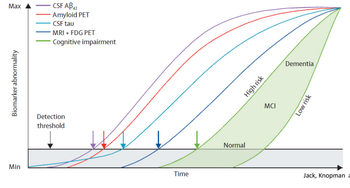

Figure: Earlier Cliff Jack model, demonstrating the progression of Amyloid (A), Tau (T), and Neurodegeneration via MRI/PET and Cognitive Impairment (N).

OBJECTIVE

Given only PET image summaries and cognitive test measures from a preclinical AD cohort, derive a summary measure that is correlated with or predictive of tau measurements.

CONTRIBUTION

Propose a formulation for analysis where a specialized measure derived only using beta-amyloid loads and longitudinal cognitive trends is meaningful: when utilized within an appropriate statistical setup uniformly improves correlations with tau measures in our heterogeneous preclinical AD cohort. We make three contributions.

- **Statistical association modeling using stochastic process**
  We use an extension of Gaussian process, a non-parametric stochastic process, that estimates a distribution of a function with a family of Gaussian distributions, to fit a regression model between cognitive scores and amyloid loads (high dimensionality is permitted): given individuals' cognitive scores and the population's cognitive scores and amyloid measures, we estimate the posterior distribution parameters (the expected individuals' amyloid measure and variance) by using families of Gaussian distribution.
- **Estimate individuals' "abnormality" or "outlierness" using deviation summary statistics**
  We leverage the idea of normative modeling on the estimated Gaussian process: this yields a summary score of individuals' *deviation* from the ``normative'' distribution comprised of the population's associations between amyloid measures and cognitive trends.
- **Uniformly improved correlation of estimated summary scores over baselines (i.e., amyloid or cognitive trends alone)**
  We found that such deviation scores always improve correlations between tau and region-wise or global PiB measures.

BACKGROUND

Gaussian Process (GP)

Given a collection of random variables, $f(x_1), \ldots, f(x_m)$, indexed by $x_1, \ldots, x_m \in X$, Gaussian Process (GP) is specified by a mean function $m(x)$ and a covariance function $k(x, x')$,

$$\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_m) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_m) \\ \vdots & \ddots & \vdots \\ k(x_m, x_1) & \cdots & k(x_m, x_m) \end{bmatrix} \right) \tag{1}$$

where $f(\cdot) \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot))$. The training set is $S = \{(x^{(i)}, y^{(i)})\}_{i=1}^{n} = X \times Y$ of i.i.d. samples where $x^{(i)} \in \mathbb{R}^{d_1}, y^{(i)} \in \mathbb{R}^{d_2}$, and the evaluation (query) set is $T = \{(x_*^{(i)}, y_*^{(i)})\}_{i=1}^{m} = X_* \times Y_*$ where $x_*^{(i)} \in X_* \subseteq \mathbb{R}^{d_1}, y_*^{(i)} \in Y_* \subseteq \mathbb{R}^{d_2}$ with a zero-mean GP prior $f(\cdot) = \mathcal{GP}(\mathbf{0}, k(\cdot, \cdot))$. From the generative modeling perspective, GP assumes $y^{(i)} = f(x^{(i)}) + \epsilon^{(i)}, i = 1, \ldots m$ where $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$. Thus, by the Gauss–Markov theorem, the conditional distribution of $\mathbf{Y}_*|\mathbf{Y}, \mathbf{X}, \mathbf{X}_*$ also follows the multivariate Gaussian distribution:

$$\mathbf{Y}_*|\mathbf{Y}, \mathbf{X}, \mathbf{X}_* \sim \mathcal{N}(\mu_i^*, \Sigma^*) \quad \text{where} \quad \mu_i^* = K(X_*, X)(K(X, X) + \sigma^2 I)^{-1}\mathbf{Y}, \tag{2}$$

$$\Sigma^* = K(X_*, X_*) + \sigma^2 I - K(X_*, X)(K(X, X) + \sigma^2 I)^{-1}K(X, X_*) \tag{3}$$

where

$$K(X, X') \in \mathbb{R}^{|X| \times |X'|}, \quad K(X, X')_{ij} = k(x^{(i)}, x'^{(j)}).$$

In summary, GP updates the distribution of functions $f(\cdot)$ and encodes the uncertainty in its covariance matrix.

Normative Modeling Scores (NMS)

Given an individual i, for each ROI j of 18 (9 bilateral) PiB ROIs, we model the cognitive scores of an individual i predictors $x^{(i)} \in \mathbb{R}^{18}$ and his/her amyloid as $y^{(ij)} \in \mathbb{R}$, and $f(\cdot) : \mathbb{R}^{18} \mapsto \mathbb{R}$. Then, we model the training set $S = \{x^{(k)}, y^{(k)}\}_{k \neq i}$ as population's cognitive scores and amyloid measures $X \times Y$ and evaluation set as singleton $T = \{(x^{(i)}, y^{(ij)})\} = X_* \times Y_*$ as individual i's predictor and response measure at ROI j from amyloid scan. GP first computes $K(\cdot, \cdot)$ from S, then the conditional distribution of $y^{(ij)}$ given $x^{(i)}$ and S, can be estimated by a normal distribution through mean μ_i and variance σ_i^2.

GP regression takes cognition scores $x^{(i)}$ and S to estimate the expected value (μ_i) and variance (σ_i^2) of the target PiB value in the region. Then, the NMS score is calculated by computing the absolute difference between the response $y^{(ij)} \in \mathbb{R}$ and mean parameter $\mu_i \in \mathbb{R}$ from GP, normalized by the sum of individuals' variance $\sigma_i^2 \in \mathbb{R}$, also from GP, and data variance $\sigma_{data}^2 \in \mathbb{R}$. The form is

$$NMS = \frac{|y^{(ij)} - \mu_i|}{\sqrt{\sigma_i^2 + \sigma_{data}^2}}. \tag{4}$$

The normalization takes into account the group level variance as well as individual level variance estimated by GP and measures the level of deviance from the cohort at the individual level.

DATASET

1. **The Wisconsin Registry for Alzheimer's Prevention (WRAP)**. A large battery of neuropsychological tests that includes cognition, lifestyle, physical activity, biomarkers, genetics, and metabolomics is given to all subjects ($N = 156$). In this experiment, the longitudinal cognitive scores consists of 16 pairs (current measurement: "intercept" and progression trajectory: "slope") of psychological scores.
2. **Pittsburgh compound B radiotracer PET image scans (PiB)**. PiB provides scans to image beta-amyloid plaques in neuronal tissue, which is widely used for Alzheimer's disease. ($N = 83$) patients were included. In this experiment, PiB observations consist of 18 ROIs (9 bilateral AD-related ROIs).
3. **Cerebrospinal fluids collection through lumbar punctures (CSF)**. In this experiment, we used hTau and pTau. ($N = 43$) out of 83 participants also have CSF features available.

METHOD

Using leave-one-out cross validation, for each participant:

1. For each PiB ROI response, 16 pairs of cognition scores serve as predictors in the GP regression model.
2. The NMS is calculated by taking the normalized absolute difference between the true PiB response and predicted response from the GP model.
3. These scores are then evaluated via their correlations with CSF, compared to a baseline of correlation between the PiB score and CSF directly.

RESULTS

Figure 2 and table 1 showed our experiment results. Table 1 shows NMS uniformly improved the correlation with tau measures compared with PiB measure. Such improvement suggests that modeling individual level of deviation from the cohort using two modalities appears to be beneficial for estimating a correlated predictive signal for tau. Figure 2 shows ROI Precuneus L between NMS and tau measurement (hTau and pTau separately).

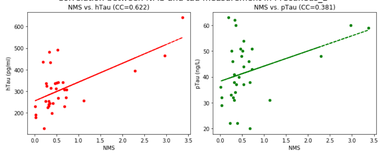

Figure: Scatter plot from region Precuneus L between NMS and tau measurement (hTau and pTau separately). CC, correlation coefficients. Best fit linear regression is shown in dash

	pTau		hTau	
	NMS	PiB	NMS	PiB
Angular_L	0.300	0.284	0.540	0.400
Angular_R	**0.361**	0.213	0.578	0.324
Cingulum_Ant_L	**0.378**	0.148	**0.635**	0.281
Cingulum_Ant_R	**0.360**	0.302	**0.611**	0.386
Cingulum_Post_L	0.227	0.313	0.467	0.397
Cingulum_Post_R	0.236	0.221	0.439	0.337
Frontal_Med_Orb_L	0.311	0.131	0.598	0.272
Frontal_Med_Orb_R	0.327	0.278	0.599	0.359
Precuneus_L	**0.381**	0.295	**0.622**	0.407
Precuneus_R	0.345	0.330	0.596	0.432
SupraMarginal_L	0.335	0.172	0.566	0.265
SupraMarginal_R	0.290	0.209	0.457	0.285
Temporal_Mid_L	**0.368**	0.210	0.497	0.327
Temporal_Mid_R	0.249	0.281	0.506	0.332
Temporal_Sup_L	0.329	0.125	0.464	0.227
Temporal_Sup_R	0.280	0.172	0.489	0.299
Precentral_L	**0.357**	0.098	0.461	0.220
Precentral_R	0.179	0.246	0.374	0.352

Table: Correlation comparisons between NMS and PiB (baseline) over tau measurement under 18 brain regions. Red shows higher correlation for each tau measurement. Bold shows relatively high correlation (> 0.35 for pTau; > 0.6 for hTau)

CONCLUSIONS

Our proposed method improves estimates correlated to tau measure given cognitive scores and beta-amyloid measure with a preclinical AD cohort.

NMS can serve as tau measurement estimates to address the time delay between amyloid abnormality and tau abnormality -- and may provide clinically useful interventions and/or enrollment in clinical trials designed for subjects at risk.

ACKNOWLEDGEMENTS

This research was supported by NIH grants R01 AG062336, R01 AG021155, R01 AG062285, and the UW-Madison Alzheimer's Disease Research Center (UW ADRC AG033514).