

## Principal Investigator

**First Name:** Paul  
**Last Name:** Schulz  
**Degree:** MD  
**Primary Affiliation:** University of Texas Health Science Center Houston  
**E-mail:** [cmkfarrell@gmail.com](mailto:cmkfarrell@gmail.com)  
**Phone number:** 713-486-0527  
**Address:** 1941 East Road Suite 4358 Houston, Texas 77054  
6431 Fannin Street  
**City:** Houston  
**State or Province:** Texas  
**Zip or Postal Code:** 77030  
**Country:** USA

## General Information

### Key Personnel (in addition to PI):

**First Name:** Christine  
**Last name:** Farrell  
**Degree:** PhD  
**Primary Affiliation:** UTHealth Science Center Houston  
**SCOPUS ID:**

**First Name:** Xiaoqian  
**Last name:** Jiang  
**Degree:** PhD  
**Primary Affiliation:** UTHealth Science Center Houston  
**SCOPUS ID:**

**First Name:** Rosemarie  
**Last name:** Schmandt  
**Degree:** PhD  
**Primary Affiliation:** UTHealth Science Center Houston  
**SCOPUS ID:**

**First Name:** Yejin  
**Last name:** Kim  
**Degree:** PhD  
**Primary Affiliation:** University of Texas Health Science Center at Houston  
**SCOPUS ID:**

**First Name:** Dulin  
**Last name:** Wang  
**Degree:** PhD Student  
**Primary Affiliation:** UTHealth Science Center at Houston  
**SCOPUS ID:**

**Are external grants or funds being used to support this research?:** No external grants or funds are being used to support this research.

**How did you learn about the YODA Project?:** Colleague

## Conflict of Interest

[https://yoda.yale.edu/system/files/2020\\_05\\_22\\_13\\_30\\_52.pdf](https://yoda.yale.edu/system/files/2020_05_22_13_30_52.pdf)  
[https://yoda.yale.edu/system/files/2020\\_05\\_22\\_13\\_31\\_21.pdf](https://yoda.yale.edu/system/files/2020_05_22_13_31_21.pdf)  
[https://yoda.yale.edu/system/files/2020\\_05\\_22\\_13\\_31\\_46.pdf](https://yoda.yale.edu/system/files/2020_05_22_13_31_46.pdf)  
[https://yoda.yale.edu/system/files/2020\\_05\\_22\\_13\\_32\\_19.pdf](https://yoda.yale.edu/system/files/2020_05_22_13_32_19.pdf)  
[https://yoda.yale.edu/system/files/yoda\\_project\\_coi\\_form\\_for\\_data\\_requestors\\_2019\\_y.\\_kim.pdf](https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_y._kim.pdf)  
[https://yoda.yale.edu/system/files/yoda\\_project\\_coi\\_form\\_for\\_data\\_requestors\\_2020.pdf](https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2020.pdf)

## Certification

**Certification:** All information is complete; I (PI) am responsible for the research; data will not be used to support litigious/commercial aims.

**Data Use Agreement Training:** As the Principal Investigator of this study, I certify that I have completed the YODA Project Data Use Agreement Training

1. [NCT00253214 - GAL-INT-10 - Placebo-Controlled Evaluation of Galantamine in the Treatment of Alzheimer's Disease: Safety and Efficacy of a Controlled-Release Formulation](#)
2. [NCT00236574 - CR003145 // GAL-INT-11 - A Randomized Double Blind Placebo-Controlled Trial to Evaluate the Efficacy and Safety of Galantamine in Patients With Mild Cognitive Impairment \(MCI\) Clinically at Risk for Development of Clinically Probable Alzheimer's Disease](#)
3. [NCT00236431 - GAL-INT-18 - A Randomized Double-Blind Placebo-Controlled Trial to Evaluate the Efficacy and Safety of Galantamine in Patients With Mild Cognitive Impairment \(MCI\) Clinically at Risk for Development of Clinically Probable Alzheimer's Disease](#)
4. [NCT00679627 - GALALZ3005 - A Randomized, Double-Blind, Placebo-controlled Trial of Long-term \(2-year\) Treatment of Galantamine in Mild to Moderately-Severe Alzheimer's Disease](#)
5. [NCT00216593 - GAL-ALZ-302 \(PMID # 19042161-CR003940\) - Treatment of Severe Alzheimer's Disease in a Residential Home, Nursing Home, or Geriatric Residential Setting: Evaluation of Efficacy and Safety of Galantamine Hydrobromide in a Randomised, Doubleblind, Placebo-Controlled Study](#)
6. [GAL-USA-10 - Placebo-controlled evaluation of galantamine in the treatment of Alzheimer's disease: Evaluation of safety and efficacy under a slow titration regimen](#)
7. [NCT00575055 - ELN115727-302 - A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Efficacy and Safety Trial of Bapineuzumab \(AAB-001, ELN115727\) In Patients With Mild to Moderate Alzheimer's Disease Who Are Apolipoprotein E4 Carriers](#)
8. [NCT00574132 - ELN115727-301 - A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Efficacy and Safety Trial of Bapineuzumab \(AAB-001, ELN115727\) In Patients With Mild to Moderate Alzheimer's Disease Who Are Apolipoprotein E4 Non- Carriers](#)

**What type of data are you looking for?:** Individual Participant-Level Data, which includes Full CSR and all supporting documentation

## Research Proposal

### Project Title

Identification of biomarkers associated with Alzheimer's disease progression that correlate with responses to medications.

### Narrative Summary:

Alzheimer's disease (AD) manifests with strikingly different ages and sites of onset, progresses at different rates and in different anatomic directions, and has varied responses to treatment<sup>1</sup>. Disease-modifying AD clinical trials have been disappointing; however, detailed analyses often show a subset of patients with improved outcomes<sup>2–5</sup>. We hypothesize that AD is heterogeneous and only subsets of patients may respond to each treatment. Using computational analysis, we propose to identify biomarkers that correlate with the four critical clinical variables (age and site of onset, rate and direction of progression) and to test whether those markers correlate with responses to AD medications.

### **Scientific Abstract:**

**Background:** Past AD trials have failed to show overall benefit of treatment, yet subsets of patients have responded. We hypothesize that AD heterogeneity is responsible for variable treatment outcomes. Our goal is to identify biomarkers that define subsets of AD patients who may respond to therapies. Our long-term goal is to be able to identify more homogeneous AD patient groups, using biomarkers, for future clinical trials.

**Objective:** We will use advanced computational analysis techniques to identify biomarkers associated with AD onset and progression and derive patient phenotype clusters for correlation with responses to AD treatments.

**Study Design:** Available variables from past clinical trials will be analyzed to identify informative biomarkers, which will then be tested for correlation with response to treatment.

**Participants:** To most accurately model real world circumstances, all trial participants will be included.

**Main Outcome Measure(s):** Biomarkers associated with disease progression and associated with responses to treatments.

**Statistical Analysis:** Bivariate analyses and non-linear regression analyses will be used to screen for informative biomarkers versus AD variables. Using advanced machine learning, we will construct patient phenotype clusters. A combination of phenotype clusters, drug characteristics, and patient response rates will be used to develop an end-to-end efficacy prediction algorithm for drug response in AD. We will mitigate the lack of drug response data in AD models using cold-start algorithms based on latent factor modeling.

### **Brief Project Background and Statement of Project Significance:**

AD affects over 5 million Americans<sup>6</sup>. One of the most challenging aspects of AD is its heterogeneity, presenting at widely different ages and sites of onset, and progressing at different rates and in different anatomic directions<sup>1</sup>. A particularly frustrating aspect of AD is its variable responses to treatment<sup>1</sup>. To date, all disease-modifying clinical trials have failed to produce a uniform beneficial outcome<sup>2,3</sup>. However, detailed post-hoc analyses have often shown that subpopulations of treated patients have improved outcomes<sup>4,5</sup>.

Due to the high prevalence of AD and the lack of good, available treatment options, there is a great need to ascertain why AD is so heterogeneous and why only some subsets of patients respond to treatments. We hypothesize that this heterogeneity is due to differences in unique biomarkers between patients<sup>7</sup>. These may include differing demographic, biologic and environmental variables. Our goal is to identify biomarkers that correlate with differential responses to treatments. Treatments could then be “personalized” according to biomarker profiles.

We propose to study biomarkers available from previous AD trials to identify those that correlate with the natural history of AD, using bivariate and non-linear regression analyses. Access to vast clinical trial datasets will allow us to identify biomarkers that have an important association with critical, clinical patient variables.

Exploration of past clinical trial data will allow us to derive unique phenotype clusters and test whether they predict responses to medications using an end to end prediction algorithm. Our long-term goal is to use biomarkers to predict a patient’s response to treatment, which would allow us to enroll fewer patients in clinical trials and have greater power to detect responses.

Identification of factors that contribute to disease progression could provide clinicians with valuable diagnostic and treatment tools upon initial encounter of a patient or throughout a patient’s disease course. Furthermore, the creation of phenotype clusters could allow clinicians to recommend specific clinical trials for patients based on their phenotype to maximize benefit to the patient and prevent unnecessary treatment. Prediction of drug response rates could allow researchers to develop clinical trial protocols to maximize treatment effect, while minimizing the monetary burden of clinical trials due to large enrollment needed to achieve statistical power. These results will also potentially increase the success of future clinical trials for AD drug treatments, ultimately providing new opportunities for AD drug development.

### **Specific Aims of the Project:**

Aim 1: Identify biomarkers associated with disease onset and progression.

We will use bivariate and non-linear regression analyses to identify biomarkers associated with AD onset and progression. Depending on available data, these will include demographic, cognitive, biological, imaging, and environmental variables.

Aim 2: Establish phenotypes of AD patients.

We hypothesize that the biomarkers identified above will allow us to identify more homogeneous groups of AD patients who share biomarker profiles. We propose to identify unique phenotype clusters using tensor factorization. These phenotypes would contain distinct characteristics that could be applied to future clinical trial patient selection.

Aim 3: Test for correlations between derived AD phenotypes and drug responses to standard and investigational medications.

Using advanced machine learning, we will test for correlations between AD phenotypes and response rates to standard and investigational medications. Our long-term goal is to identify biomarkers that classify AD patient into groups that are likely to respond similarly to treatments, but differently from other AD patient subsets. Such results will allow for better patient selection for trials in terms of likely responders and non-responders.

### **What is the purpose of the analysis being proposed? Please select all that apply.**

New research question to examine treatment effectiveness on secondary endpoints and/or within subgroup populations

Confirm or validate previously conducted research on treatment effectiveness

Preliminary research to be used as part of a grant proposal

Participant-level data meta-analysis

Participant-level data meta-analysis pooling data from YODA Project with other additional data sources

Research on comparison group

Research on clinical prediction or risk prediction

### **Research Methods**

#### **Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:**

Patient-level data, including cognitive testing, laboratory results, concomitant medications, exploratory biological sample data, raw imaging files, and imaging results for all subjects that screened and/or enrolled in past clinical trials through the YODA Project for mild cognitive impairment, early, mild, or moderate AD is requested. Placebo and treatment group data will be assessed as available.

Full CSR and all supporting documents, particularly those detailing the drug chemistry and structure is also requested. To enhance the efficacy of our prediction algorithm, standard drug chemistry and structure will be incorporated into the analysis from the DrugBank database or similar. DrugBank is a public database with over 13,000 drug target information.

Incomplete data will be excluded. In order to incorporate the most real world data to inform disease progression and form phenotype clusters, no patient-level inclusion or exclusion criteria (ex: age, medical history, medications) will be placed on data to be analyzed. Potential confounders, however, include the different inclusion/exclusion criteria applied to each clinical trial dataset at the time it was conducted.

#### **Main Outcome Measure and how it will be categorized/defined for your study:**

Aim 1: Biomarkers associated with Clinical Variables

The main outcome measure will be biomarkers that correlate with the AD clinical variables of age and site of onset, and rate and anatomic direction of progression. Progression will be defined by serial cognitive testing results or sequential imaging findings. Clinical onset will be established by cognitive testing, genomics, or imaging findings.

Aim 2: AD Phenotypes

The main outcome measure will be phenotype clusters established based on decreased variance within groups versus between groups. Successful phenotypes will be those that group patients with distinct characteristics and

remain significant in the validation cohort.

### Aim 3: Correlations between AD phenotypes and Medication Responses

The main outcome measure will be positive correlations between AD phenotypes and responses to standard and experimental medications. We propose to apply advanced computational methods to curated standard and investigational drug information. Drug effect size will be calculated using area under the ROC curve.

### **Main Predictor/Independent Variable and how it will be categorized/defined for your study:**

#### Aim 1: Biomarkers associated with Clinical Variables

The main predictor will be biomarkers associated with age and site of onset, and rate and anatomic direction of progression. Significant biomarkers will be defined as any covariate with an association of  $p < 0.05$ .

#### Aim 2: AD Phenotypes

The main predictor is the characteristics used to establish the phenotypes. These characteristics could include differing demographics, cognitive function, genomics, epigenomics, transcriptomics, proteomics, metabolomics, MRI and PET imaging findings, environmental variables, and medical history and concomitant medications.

#### Aim 3: Correlations between AD phenotypes and Medication Responses

We hypothesize that drug response should be more similar within phenotype clusters than between clusters. The main predictors will be biomarkers that distinguish AD phenotypes that then correlate with responses to standard or investigational medications.

### **Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:**

Variables of interest include: gender; ethnicity; initial and serial cognitive testing; laboratory results: exploratory genomic data, epigenomics, transcriptomics, proteomics, metabolomics; MRI and PET imaging raw data and results; and environmental factors, when available. Cognitive testing conducted in each clinical trial may differ, therefore classification of results may be necessary for comparison across trials.

Drug target information for investigational medications from requested studies and standard and discovery-phase drugs from DrugBank will be incorporated into the prediction algorithm to enhance the predictive value.

Laboratory results may not be available from all datasets and may require evaluations on different sets of normal values prior to comparison across trials.

Ideally raw imaging files would be available for volumetric and serial evaluation, however, imaging reports would also enhance the robustness of the data.

Environmental factors could be highly variable in type and collection method, and will most likely require classification prior to comparison across datasets.

### **Statistical Analysis Plan:**

Bivariate and non-linear regression analyses will be used in Aim 1 to identify biomarkers associated with clinical disease variables throughout the AD spectrum. Bivariate analyses will utilize Kendall correlation, Kruskal-Wallis one way rank ANOVA, and the Mann-Whitney U test, as appropriate. Independent predictors of age and site of onset, and rate and anatomic direction of progression ( $p < 0.05$ ) will be further assessed through non-linear regression analyses.

Our hypothesis mandates the use of either large treatment groups, or the combination of data from multiple datasets, in order to achieve statistical power to predict unique biomarker profiles. However, due to variations in acquisition methods and observations, biomarker predictions or treatment effects may be muted. Using a multi-modal, computational phenotyping method, nonnegative tensor factorization<sup>8,9</sup>, we can analyze multisite datasets across past clinical trials in order to maximize our sample population while mitigating the risks associated with noisy data.

Significant biomarkers identified in Aim 1, will provide the insights to facilitate Aim 2. The analyses for Aim 2 will

include advanced machine learning, coupled nonnegative tensor factorization, to group patients into phenotype clusters. These patient biomarker profiles would be based on variables that are more similar within the cluster, than between clusters. Biological, imaging, demographic, and drug-effect variables will be incorporated in order to derive phenotype clusters. This multi-modal, computational method will enhance the robustness of the phenotypes. Phenotypes will be filtered to remove less discriminative ones based on statistical significance ( $p > 0.05$ ), and in order to establish phenotypes with distinct characteristics. This method will also provide the membership values of each variable, in order to determine the contribution of each variable to the derived phenotype. Phenotypes will be derived using a small test cohort of the population. Discriminative ability of each derived phenotype will be verified using the remaining population, validation cohort. The resulting phenotypes will be combined with additional data for use in Aim 3.

A similar study utilizing data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database was previously performed by our group to determine phenotypes<sup>7</sup>. This analysis, however, was limited to the variables and cases available in the ADNI database. While certain limitations still exist, the use of large, well-characterized, diverse datasets of patients from past trials will enhance our ability to detect and identify relevant biomarkers, create phenotype clusters, and develop a predictive algorithm for drug response.

Patient response rates to standard and investigational medications in models of AD and drug molecular features, along with the established phenotypes will be aggregated for use in Aim 3. The DrugBank database of drug target information will be uploaded to the YODA platform and also included in analyses. We will then test the ability of individual phenotype clusters to predict the response to AD treatments and experimental drugs using an end-to-end prediction algorithm. Individuals that are naïve to the treatment being assessed, non-responders to treatments, and initial responders to non-FDA approved medications will be used for prediction. Effect size will be calculated using area under the ROC curve in each placebo/treatment pair for each clinical trial dataset in each phenotype group. The same analysis will be completed through all derived phenotype clusters. This type of analysis will allow for prediction of improvement or worsening upon exposure to a drug of interest. The addition of targets from DrugBank will allow for prediction of response to discovery-phase drugs as well. We will mitigate the lack of drug response data in AD models using cold-start algorithms based on latent factor modeling.

Software Used:

R

#### **Project Timeline:**

The projected time for this project is 12 months. The anticipated start date is August 2020, but will be contingent upon data acquisition from all sources. Analysis completion is projected to take nine months: Aim 1 analysis will be from months 1-2; months 2-6 will allow for completion of analyses for Aim 2; and Aim 3 analyses will be completed in months 6-9. Specifically, any classification or consolidation of variables required for bivariate or multivariate analyses in Aim 1 will be conducted during month 1. These variables will then be available in month 2 to complete Aim 1 and begin phenotyping analysis in Aim 2. Selection and validation of phenotypes will be conducted in month 4-6. Preparation for the predictive algorithm, including acquisition of any external drug data, will be completed in month 6, and the model will be tested in months 7-9.

Based on available data, multiple manuscripts may be written. The first manuscript will be drafted during month ten and submitted for publication, with subsequent manuscripts to follow. Results will be reported back to the YODA Project at the time of manuscript submission. Results may also be incorporated into future grant proposals for biomarker-related studies.

#### **Dissemination Plan:**

The results of this project are potentially broad with clinical and research applications. Depending on the number of variables available for analysis, the first manuscript will highlight the variables associated with disease progression. This paper will be targeted for a clinical audience, as it has the greatest implications for initial patient treatment and modifications to treatments over time. Therefore, we will focus on Alzheimer's & Dementia: Translational Research & Clinical Interventions (TRCI) or similar journals. This data may also be incorporated into biomarker-driven projects or grant proposals.

An additional manuscript will detail the derivation of phenotype clusters and the ability to predict drug response. Clinicians could value this topic as it applies to their recommendation of patients to clinical trials and treatment with existing medications. Researchers and clinical trialists could utilize these predictive tools to enhance patient selection and reduce monetary burdens of conducting a clinical trial. Due to the clinical and research ramifications

of this topic, this manuscript will be submitted to JAMA Neurology or similar journals, as its audience encompasses both sectors. All manuscripts will be shared with the YODA Project at the time of submission.

### **Bibliography:**

1. Lam, B., Masellis, M., Freedman, M., Stuss, D. T. & Black, S. E. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimers Res Ther* 5, 1 (2013).
2. Cummings, J. L., Morstorf, T. & Zhong, K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther* 6, 37 (2014).
3. Liu, P.-P., Xie, Y., Meng, X.-Y. & Kang, J.-S. History and progress of hypotheses and clinical trials for Alzheimer's disease. *Sig Transduct Target Ther* 4, 29 (2019).
4. Howard, R. & Liu, K. Y. Questions EMERGE as Biogen claims aducanumab turnaround. *Nat Rev Neurol* 16, 63–64 (2020).
5. Siemers, E. R. et al. Phase 3 solanezumab trials: Secondary outcomes in mild Alzheimer's disease patients. *Alzheimer's & Dementia* 12, 110–120 (2016).
6. 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* 16, 391–460 (2020).
7. Alzheimer's Disease Neuroimaging Initiative et al. Multimodal Phenotyping of Alzheimer's Disease with Longitudinal Magnetic Resonance Imaging and Cognitive Function Data. *Sci Rep* 10, 5527 (2020).
8. Ho, J. C. et al. Limestone: high-throughput candidate phenotype generation via tensor factorization. *J Biomed Inform* 52, 199–211 (2014).
9. Choi, J., Kim, Y., Kim, H.-S., Choi, I. Y. & Yu, H. Phenotyping of Korean patients with better-than-expected efficacy of moderate-intensity statins using tensor factorization. *PLoS ONE* 13, e0197518 (2018).