Principal Investigator

First Name: Hongtu
Last Name: Zhu
Degree: PhD
Primary Affiliation: UNC-Chapel Hill
E-mail: htzhu@email.unc.edu
State or Province: North Carolina
Country: United States

General Information

Key Personnel (other than PI):
First Name: Tengfei
Last name: Li
Degree: PhD
Primary Affiliation: UNC-Chapel Hill
SCOPUS ID:
Requires Data Access? Yes

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/wp-content/uploads/2024/03/SV_57KskaKADT3U9Aq-R_36TFS3Qg3Ysj0l3.pdf
https://yoda.yale.edu/wp-content/uploads/2024/03/COI_Li.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. NCT00236574 - 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
2. NCT00253214 - Placebo-Controlled Evaluation of Galantamine in the Treatment of Alzheimer's Disease: Safety and Efficacy of a Controlled-Release Formulation

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

Research Proposal

Project Title

Deciphering AD Etiology: A Synthesis of Trial and Observational Data

Narrative Summary:

Deciphering AD Etiology: A Synthesis of Trial and Observational Data
This integrated approach is proposed to fill critical knowledge gaps and catalyze the development of targeted, effective interventions to combat the AD/ADRD disease. Our research endeavor seeks to meticulously merge data from influential clinical trials with comprehensive observational cohorts such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the National Alzheimer's Coordinating Center (NACC). Our goal is to unravel the intricate effects of drug treatments while examining the role of established risk factors, utilizing cutting-edge causal inference algorithms and predictive modeling.

By leveraging advanced statistical models and machine learning techniques, we intend to dissect the complex interplay between genetic predispositions, lifestyle factors, and therapeutic agents, thereby offering groundbreaking insights into AD causality and progression. We will create a robust analytical framework capable of identifying and clarifying the causal effects of pharmacological interventions and diverse health/risk factors (such as smoking and education) in the pathogenesis of Alzheimer's Disease (AD).

Scientific Abstract:

Background: This project proposed a novel approach in the comprehensive analysis of clinical trial data in conjunction with observational study data from ADNI and NACC to investigate the longitudinal (dynamic) causal effects of AD-related drugs and other risk factors on the progression of the disease.
Objective: Our specific objectives are three-fold: firstly, to identify the direct and indirect longitudinal effects of these drugs on cognitive function; secondly, to merge varying clinical data to investigate the drugs' causal effects; and thirdly, to integrate clinical with observational data to further elucidate these effects.
Study Design: To meet these aims, we will develop a dynamic causal acyclic graph, a federated learning framework, and a data combination framework, respectively. Additionally, we will craft companion software to enhance the utility of the statistical tools developed.
Participants: Participants should include those who have at least one observation of drug treatment and the cognitive function assessment.
Primary and Secondary Outcome Measures: Our primary outcome measures include longitudinal observations of cognitive assessment scores such as ADAS-Cog, CDR-SB, MMSE, and our secondary outcome measures include longitudinal observations of imaging measures and functional measures such as MRI, ECG.
Statistical Analysis Plan: Statistical Analysis plans include applying mediation analysis within a counterfactual inference framework, propensity score matching, and generalized estimating equations in modeling, amongst others. Through this multifaceted approach, our research strives to uncover the causal relationships between AD-related drugs, risk factors, and cognitive health, potentially leading to refined treatment modalities, novel therapeutic targets, and predictive indicators for disease progression, thereby revolutionizing the approach to AD management and care.

Brief Project Background and Statement of Project Significance:

Alzheimer's Disease (AD) represents a multifaceted challenge in medical research, marked by its complex pathophysiology and the significant impact it has on patients, caregivers, and healthcare systems globally. With a rising prevalence and no definitive cure, there is an urgent need to deepen our understanding of the disease's progression and the impact of therapeutic interventions. The cornerstone of this project is the comprehensive analysis of clinical trial data in conjunction with observational study data to investigate the longitudinal (dynamic) causal effects of AD-related drugs

and other risk factors on the progression of the disease.

The project aims to forge a comprehensive understanding of Alzheimer’s Disease (AD) by amalgamating clinical data with observational datasets, such as those from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the National Alzheimer's Coordinating Center (NACC). By integrating these data with existing observational datasets, we will create a rich, multidimensional perspective that can account for a wide array of risk factors and their interplay in the context of AD. The overarching goal is to discern the causal effects of AD-related drugs and other risk factors on these key aspects of neurological health.

The significance of this project is manifold. Alzheimer’s Disease, a condition that progressively impairs memory and cognitive function, poses a significant challenge to healthcare systems worldwide. The complexity of its pathology and the multifactorial nature of its progression demand a multifaceted approach to research. This project addresses this need by aiming to create a comprehensive analytical framework that integrates diverse data types. The resultant insights have the potential to refine current treatment modalities, identify novel therapeutic targets, and offer predictive indicators of disease progression. The insights gleaned could revolutionize current understandings and lead to more targeted, effective interventions.

Specific Aims of the Project:

In this proposal, we identify three challenges that we face in analyzing the clinical data and the observational data:
(CH1) Identifying the longitudinal direct effect of the drugs on the cognitive function, and the indirect effect on the cognitive function through other mediators.
(CH2) Integrating of different clinical data to investigate the causal effect of AD-related drugs.
(CH3) Integrating clinical data and observational data to investigate the causal effect of the drugs.
To address these challenges, we will develop
(Aim 1) a longitudinal (dynamic) causal acyclic graph by arrows between the drugs and other factors to the cognitive function for (CH1);
(Aim 2) a federated learning framework that can integrates information across different studies for (CH2);
(Aim 3) a data combination framework that can integrates clinical trial data with the observational study data for (CH3);
Finally, we will develop companion software for all the statistical tools developed in this project.

The general aim of this research lies in unraveling the nuanced causal relationships between AD-related drugs, risk factors, and their subsequent impact on cognitive health. By meticulously integrating disparate clinical data, this study seeks to illuminate causal effects of AD-related pharmacotherapies on cognitive functions.

Study Design:

Other

Study Design Explanation:

Meta-analysis and Methodological research

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

Participant-level data meta-analysis

Meta-analysis using data from the YODA Project and other data sources

Develop or refine statistical methods

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:

Data Source: Our study will integrate data from clinical trials with existing datasets from ADNI and NACC. The clinical trial data will serve as the primary source for analyzing the efficacy and safety of AD-related treatments, while the ADNI and NACC datasets will provide auxiliary observational data to support and enhance our causal analyses.

For datasets that are publicly available, our analysis will be conducted on a university-affiliated secure server, which employs ONYEN-password encryption to ensure data security. For clinical trial datasets where direct downloading is not permitted, we will execute our analytical scripts directly on the workbench or online server provided by the clinical trial's data management team. This approach allows us to adhere to data access restrictions while still performing comprehensive analyses.

Inclusion Criterion: participants who had at least one observation of drug treatment and the cognitive function assessment.

Primary and Secondary Outcome Measure(s) and how they will be categorized/defined for your study:

Primary outcome measures: longitudinal observations of cognitive assessment scores such as ADAS-Cog, CDR-SB, MMSE.
Secondary outcome measures: longitudinal observations of imaging measures and functional measures such as MRI, ECG.

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

Treatment covariates: drug information of the participants.
Confounding covariates: demographic information.
Mediators: imaging or functional exposures.

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

All the variables has been defined in the outcome measures and the predictor variables.

Statistical Analysis Plan:

With these clinical trial datasets, the first step is the data-processing. Our team has extensive experiences in the processing of imaging data. We will apply different imaging analysis tools including 3DSlicer, FSL, SPM, Freesurfer, DICCCOL, Loni, HAMMER, and DTI-ITK, to process all neuroimaging data. We will extract other phenotypes such as various neuropsychological test scores, CSF biomarkers, and clinical/functional status.

After data pre-processing, the proper longitudinal causal effect analysis of these clinical trial datasets and the observational data is a critical step for identifying the treatment effect of the drugs, so as to learn whether specific drugs provide clinical benefit to people with AD. Specifically, we will achieve the following:

(a) Aim 1: Construct a series of longitudinal causal acyclic graphs that delineate the progression and interplay of drug interventions, intermediary biological or behavioral factors, and their cumulative impact on cognitive functions over time.
• Propose structural equation modeling (SEM) that accommodates both drug, imaging and clinical covariates to quantify direct and indirect effects, with cognitive scores as the response variable and drug dosages, usage duration, and mediator variables as predictors. The indirect effects will be quantified through the drug effect on the mediator variable such as imaging covariate and intermediary factors.
• Incorporate latent variables in the model to account for unobserved heterogeneity and to capture the underlying constructs that may influence both mediator and outcome variables.
• Conduct mediation analysis within a counterfactual inference framework to unveil the mechanisms through which drug interventions exert their influence.
• Establish and test hypotheses regarding mediator variables that may serve as conduits for the treatment effect, such as biological markers or psychosocial factors.

(b) Aim 2: Develop a federated learning framework to aggregate information from various clinical trial data sources without compromising data privacy to enhance the precision of the causal effect estimator.
• Implement meta-analysis techniques to combine the causal effect estimator of the same drug derived from different clinical trial studies. Within each study, the estimator is based on the method proposed in Aim 1.
• For the estimator of a specific clinical trial study, information from other clinical trial studies are selected if the two datasets do not have large heterogeneity. Bias and the corresponding variance are measured to account for the heterogeneity.
• Deploy distributed algorithms capable of performing model estimation in a decentralized manner, with local computations carried out within the confines of each participating entity.

(c) Aim 3: Construct a methodological framework that systematically merges data from clinical trial data with data from observational studies. The intent is to develop a comprehensive evidence base that supports broader and more nuanced clinical insights.
• Utilize propensity score matching to create a synthetic control group from observational data that is statistically similar to the treatment groups in clinical trial, thereby reducing confounding biases.
• Apply generalized estimating equations (GEE) to model the mean response based on both experimental and observational data, while appropriately accounting for the correlation of repeated measures within subjects across different study designs.
• Integrate findings from clinical trial studies and observational studies using a sequential, layered approach that first analyzes clinical trial data for efficacy under controlled conditions, followed by observational data for effectiveness in real-world settings.

Software Used:

Python

Project Timeline:

The proposed project timeline is: Year 1: Aim 1; Year 2: Aim 2; Year 3: Aim 3. The algorithms and tools developed in this project will be released to the public for free download every year.

Project start day: July 1, 2024.
For Aim 1:
July-September 2024: Literature review and initial modeling framework development.
October-December 2024: Real data analysis of the longitudinal clinical trial data.
January-March 2025: Draft the methodology paper and the real data analysis paper.
April-June 2025: Polish the papers and submit the two papers, and report the corresponding results back to the YODA project.

For Aim 2:

July-September 2025: Literature review and initial modeling framework development.
October-December 2025: Real data analysis of multiple longitudinal clinical trial data.
January-March 2026: Draft the methodology paper and the real data analysis paper.
April-June 2026: Polish the papers and submit the two papers, and report the corresponding results back to the YODA project.

For Aim 3:
July-September 2026: Literature review and initial modeling framework development.
October-December 2026: Real data analysis of the longitudinal clinical trial data and the observational study data.
January-March 2027: Draft the methodology paper and the real data analysis paper.
April-June 2027: Polish the papers and submit the two papers, and report the corresponding results back to the YODA project.

Dissemination Plan:

We plan to write 3 papers in this project designed for the three aims, respectively. The targeted audiences are clinical researchers, practitioners, and statisticians.
The potentially suitable journals including Biometrika, Bioinformatics, Biostatistics, Biometrics, Journal of the American Statistical Association, Statistical Methods in Medical Research, Statistical Methods in Medical Research.

Bibliography:

Bibliography

[1] Colnet, B., Mayer, I., Chen, G., Dieng, A., Li, R., Varoquaux, G., ... & Yang, S. (2024). Causal inference methods for combining randomized trials and observational studies: a review. Statistical science, 39(1), 165-191.

[2] Cummings, J., Aisen, P. S., DuBois, B., Frölich, L., Jack, C. R., Jones, R. W., ... & Scheltens, P. (2016). Drug development in Alzheimer’s disease: the path to 2025. Alzheimer’s research & therapy, 8, 1-12.

[3] Fan, L., Mao, C., Hu, X., Zhang, S., Yang, Z., Hu, Z., ... & Xu, Y. (2020). New insights into the pathogenesis of Alzheimer’s disease. Frontiers in neurology, 10, 502538.

[4] Gómez-Gallego, M., Gómez-García, J., & Ato-Lozano, E. (2017). The mediating role of depression in the association between disability and quality of life in Alzheimer’s disease. Aging & mental health, 21(2), 163-172.

[5] Liu, Y., Li, Z., Ge, Q., Lin, N., & Xiong, M. (2019). Deep feature selection and causal analysis of Alzheimer’s disease. Frontiers in neuroscience, 13, 460841.

[6] Shi, X., Pan, Z., & Miao, W. (2023). Data integration in causal inference. Wiley Interdisciplinary Reviews: Computational Statistics, 15(1), e1581.

[7] Silva, M. V. F., Loures, C. D. M. G., Alves, L. C. V., de Souza, L. C., Borges, K. B. G., & Carvalho, M. D. G. (2019). Alzheimer's disease: risk factors and potentially protective measures. Journal of biomedical science, 26, 1-11.

[8] Wang, X., Liu, J., Hu, S. E. S., Liu, Z., Lu, H., & Liu, L. (2023). HILAMA: High-dimensional multi-omic mediation analysis with latent confounding. bioRxiv, 2023-09.

[9] Wei, K., Xue, F., Xu, Q., Yuan, Y., Zhang, Y., Qin, G., ... & Qu, A. (2024). Time-varying mediation analysis for incomplete data with application to DNA methylation study for PTSD. bioRxiv, 2024-02.

[10] Yang, B., Tao, B., Yin, Q., Chai, Z., Xu, L., Zhao, Q., & Wang, J. (2022). Associations Between Oral Health Status, Perceived Stress, and Neuropsychiatric Symptoms Among Community Individuals With Alzheimer's Disease: A Mediation Analysis. Frontiers in Aging Neuroscience, 13, 801209.

[11] Yang, S., Gao, C., Zeng, D., & Wang, X. (2023). Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation. Journal of the Royal Statistical Society Series B: Statistical Methodology, 85(3), 575-596.