

Background: There is significant potential for leveraging Artificial Intelligence and Machine Learning for early prediction of Alzheimer's and Related Dementias (ARD). Widely available datasets such as the Alzheimer's Disease Neuroimaging Project (ADNI) provide rich datasets including clinical data, cognitive assessments, imaging, genetic, and cerebral spinal fluid measurements. However, even though ARD disproportionately impacts people of color, ADNI has an underrepresentation of minoritized populations¹. Electronic health record (EHR) data proves to be a promising low-cost and more representative alternative, but there is only one widely available de-identified EHR dataset, Mimic-IV, which represents a single healthcare setting. There is a concern as to whether this dataset could be generalized to other healthcare settings. There is a need to expand representation, but the question remains as to the direction. Should we target enhancing research datasets, or more widely available and less expensive to collect medical record data? I think the answer is both. EHR data provides a low-cost, scalable, and widely available alternative option to quickly triage patients for potential referral for further evaluation such as cognitive assessments, imaging, or cerebral spinal fluid measurements. Having a more expansive EHR-based dataset of low-cost and easy-to-collect passive digital biomarkers is quite useful. However, to improve differential diagnosis (ie, distinguishing between vascular dementia and Alzheimer's Dementia), predict ARD in the very early pre-clinical phase, and understand mechanisms underlying ARD so that targeted therapies can be developed, more expansive research datasets that include biological-based biomarkers are essential. Thus, Ellison Laboratories proposes a patient-centric and community-based approach to expanding the diversity and availability of both research datasets and EHR-derived datasets, starting first with a focus on the African American population and later expanding to other minoritized populations. We propose a de-identified repository of EHR and claim data as well as the collection of additional research data that links to the ADNI project (same data collected and the same format).

Basic Information: To collect EHR data for our de-identified data repository, we will tap into and aggregate existing data application programming interfaces. Leading EHR vendors (Epic, Cerner, Athena Health) provide the ability for patients to consent to having their electronic health records shared with third-party apps. The Center for Medicare and Medicaid Services provides this option also through their Blue Button Initiative. Ellison Laboratories is currently developing an application that brings together data from disparate care settings to develop a longitudinal patient record so that patients, providers, caregivers, and insurers/case management have transparency into healthcare encounters across the healthcare continuum to better manage care. This app taps into the existing EHR and CMS data feeds. We currently have the capability of ingesting data from Epic and are working on implementing CMS and other EHR data sources. Our app will also ask additional questions to collect additional social determinants of health data. The application utilizes generative AI to provide user-friendly healthcare summaries and provides dashboards with key information such as medication, and lab values over time. We intend to allow patients to opt in to have their data de-identified for research purposes with compensation and a clear understanding of data usage details. This provides patients the opportunity to decide how their data is utilized and to also monetize their data if they so choose. We intend to take a grassroots community-based approach to promoting our application. Since

we are focused on the African American population, this will include targeting cities with a large African American population such as Atlanta, Houston, Detroit, Los Angeles, and New York. Our approach will involve engaging African American community leaders and in particular faith leaders. We will begin with Detroit as we have an initial customer in Detroit that is a small health insurer that serves underserved populations and has a high African American patient membership. This initial customer will be piloting our longitudinal patient record for their wellness initiatives. This initial customer can also assist us with patient recruitment. Detroit is also the location of our community/patient advisory board which will help us grow ties with the community. We are considering seeking Small Business in Innovative Research grant funding for this initiative.s

To collect research data, we propose a data collection initiative that establishes data collection sites in cities with a high African American population. We will leverage the same community-based approach that engages community leaders and faith leaders for participant recruitment. For research data collection we intend to start with Atlanta, as that is where our advisor, Laroy Penix is located. Dr. Laroy Penix is a neurology hospitalist and assistant professor at Emory University Medical Center (formerly Johns Hopkins, the University of Kentucky College of Medicine, and Morehouse School of Medicine). The goal is to link data to the ADNI project, so we intend to collect the same data collected with ADNI (imaging, clinical data, cognitive assessments, genetic data, and cerebral spinal fluid measurements) but wish to also collect additional social determinants of health. For our EHR and claim data collection, we intend to collect data from 8000 patients. This accounts for .1% of African American patients who suffer from ADRD². We understand that it will be far more difficult to collect the research data, and thus expect to collect data from 800 patients, which accounts for 0.01% of African American patients who suffer from ADRD in the US.

Major challenges of our plan include patient recruitment and data privacy and security. Although our CEO is certified via Hi-Trust to de-identify patient data using the expert method, and Ellison Laboratories has completed the CMS data management and security pilot program to be permitted to store CMS claims data in the cloud, we will retain IQVIA to certify our data-deidentification process via the expert method. As to the challenge of patient recruitment, as an African American-owned organization we understand the hesitancy of African Americans to participate in data collection initiatives and healthcare research. There is an unfortunate history of African Americans being harmed in these settings. To combat this, we plan to take a community-based approach to recruitment that engages community leaders, faith communities, and other events within the community for recruitment. We also intend to pursue grant funding to provide generous compensation for participants and will also provide educational materials in a variety of mediums as well as boots on the ground to explain the goals of the research study, the importance to our community, and to ascertain that participants who consent are properly informed as to the goals, benefits, and drawbacks of the study.

Utility & Rigor

To ascertain that data can be linked back to ADNI, for research data collection, we will follow the existing methodology of ADNI. There are numerous studies in the literature that utilize ADNI to predict ADRD^{3,4}. Numerous studies have also utilized electronic health-record data to predict ADRD data⁵⁻⁹. Ellison Laboratories was able to predict ADRD from the Mimic-IV EHR dataset with ~0.9 AUC, which is comparable to prior studies leveraging similar data. To determine whether an ADRD diagnosis has occurred, we are using two inclusion criteria. The first inclusion criterion includes the presence of an ICD-9 or ICD-10 code associated with ADRD in the medical record. The second inclusion criterion includes the presence of a medication utilized to treat ADRD in the medical record. We consider a patient to have an ADRD diagnosis if either criterion is met. For the ICD code criterion, we are using the ICD-9 and ICD-10 codes outlined in ^{10,11}. Like ¹¹ the medications in our inclusion criteria are rivastigmine, galantamine, donepezil, memantine, and aducanumab.

Innovation

While datasets such as Mimic-IV and ADNI have greatly advanced Data Science research efforts in the aging space, ADNI is limited due to its lack of representation of minoritized patients, and Mimic-IV is limited in that it sources data from a single medical setting. Our solution builds upon the important work of ADNI but expands it to include a currently underrepresented population (ADNI has <5% African American patients). Our work also provides the capacity to expand the use and availability of EHR and claim data for ADRD research efforts beyond Mimic-IV. Our innovative patient-centric approach in which patients are the decision maker regarding ingesting data into the system alleviates common issues related to data silos and data sharing across institutions. It also empowers the patient to be the ultimate decision-maker as to how their data is being used. We also take a hybrid approach recognizing the need for both inexpensive and scalable methods such as EHR data collection and more robust biologically based biomarker approaches that include the collection of more expensive data such as imaging, cognitive assessments, genetic data, and cerebral spinal fluid measurements.

Disproportionate Impact

The core of our approach is centered on improving upon the diversity of research and medical record datasets available for ADRD research. This includes a community-driven approach to reaching first the African American population and other minoritized populations to follow. Our efforts will first target Detroit, MI, and Atlanta, GA, and be expanded to other cities in the future. Patients will be compensated for their participation and we plan to apply for NIH grant funding to support our efforts.

Feasibility

Major challenges for a project of this nature include patient recruitment, privacy and security, and data sharing across institutions. Our patient-centric approach in which the patient is the decision maker for sharing data allows us to source data directly from EHR and CMS claim data feeds. Our proof of concept application allows patients to select various health settings for which they have visited. They are then permitted to elect which datasets they wish to share with our application (CMS claim data, Epic data, Cerner data, Athena Health data, etc). They are then routed to the data portal housing the original data to consent to data sharing. Once this step is completed, we can ingest data into our system. This can occur without additional institutional approval. Regarding privacy and security, we have completed the CMS data privacy and security pilot program which allows Ellison Labs to be one of the first vendors to store CMS claim data in the cloud. Our CEO is also certified to conduct HIPAA data de-identification via the expert method through HI-Trust. We will also certify our de-identification process using well-respected vendor Privacy Analytics/IQVIA. Regarding member recruitment, we plan to take a community-based approach and engage local community leaders and faith leaders. We also have additional credibility with the community as we are a 100% African-American-owned and majority-minority team.

We also have a highly qualified team. Ellison Laboratories is led by our CEO, Dr. Korin Reid. Korin Reid earned her Ph.D. from Georgia Institute of Technology. Since then, Dr. Reid has held a variety of positions in the healthcare IT space including Senior Data Scientist at Optum/UnitedHealthcare/MckessonHealth Solutions, and VP of Data and Innovation at Craneware. Dr. Reid has built AI solutions that have impacted 160 million patients and a third of US hospitals and was named to Forbes 30 Under 30 in 2017 for her work in this space. Dr. Reid also teaches Data Science at the University of California at Berkeley. Ellison Laboratories has a stellar group of advisors including Dr. Laroy Penix, a neurology hospitalist and assistant professor at Emory University Medical Center (formerly Johns Hopkins, the University of Kentucky College of Medicine, and Morehouse School of Medicine). We are also advised by Dr. Darryl Carter, pathologist, entrepreneur in residence at the University of Maryland (formerly Johns Hopkins), and serial entrepreneur and founder of Nora Therapeutics. Ellison Laboratories is a National Institute of Aging Startup Challenge Accelerator Finalist, Matter Brain Health Challenge Award Winner, Microsoft for Startups Accelerator Participant, and an NVIDIA Accelerator Participant. Ellison Laboratories was among the first vendors permitted to store Center for Medicare and Medicaid Services claim data in the cloud as we completed their Data Management and Security Pilot program. Our CEO, Dr. Korin Reid is certified in HIPAA healthcare data de-identification via the expert method through HI-Trust. We plan to build upon our current capabilities to assemble a high-quality multi-disciplinary research team of clinicians, researchers, and community members to carry out the goals of this project.

1. Lim, A. C. et al. Quantification of race/ethnicity representation in Alzheimer's disease neuroimaging research in the USA: a systematic review. *Commun. Med.* **3**, 101 (2023).
2. Black Americans and Alzheimer's. *Alzheimer's Disease and Dementia*
<https://www.alz.org/help-support/resources/black-americans-and-alzheimers>.
3. El-Sappagh, S., Alonso, J. M., Islam, S. M. R., Sultan, A. M. & Kwak, K. S. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Sci. Rep.* **11**, 2660 (2021).
4. Venugopalan, J., Tong, L., Hassanzadeh, H. R. & Wang, M. D. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci. Rep.* **11**, 3254 (2021).
5. Tjandra, D., Migrino, R. Q., Giordani, B. & Wiens, J. Cohort discovery and risk stratification for Alzheimer's disease: an electronic health record-based approach. *Alzheimers. Dement.* **6**, e12035 (2020).
6. Gao, X. R. et al. Explainable machine learning aggregates polygenic risk scores and electronic health records for Alzheimer's disease prediction. *Sci. Rep.* **13**, 450 (2023).
7. Xu, J. et al. Data-driven discovery of probable Alzheimer's disease and related dementia subphenotypes using electronic health records. *Learn. Health Syst.* **4**, e10246 (2020).
8. Li, Q. et al. Early prediction of Alzheimer's disease and related dementias using real-world electronic health records. *Alzheimers. Dement.* (2023) doi:10.1002/alz.12967.
9. Park, J. H. et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *bioRxiv* (2019) doi:10.1101/625582.
10. Bush, K., Wilkinson, T., Schnier, C., Nolan, J. & Sudlow, C. Definitions of dementia and the major diagnostic pathologies, UK biobank phase 1 outcomes adjudication.
https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/alg_outcome_dementia.pdf.
11. Duan, R. et al. Leverage Real-world Longitudinal Data in Large Clinical Research Networks for Alzheimer's Disease and Related Dementia (ADRD). *AMIA Annu. Symp. Proc.* **2020**, 393–401 (2020).