Executive Summary

Latinos are the largest and fastest-growing minority in the United States,[1] numbered 63 million in 2022, with nearly 60% of Mexican origin.[2,3] Latinos are disproportionately more likely than older white adults to suffer from AD/ADRD.[4] We propose to discover novel markers for early prediction of AD/ADRD using machine learning (ML) approaches, the Mexican Health and Aging Study (MHAS), and the Mexican Cognitive Aging Ancillary Study (Mex-Cog) database. Mex-Cog is an in-depth cognitive evaluation of participants in the MHAS.[5] The Mex-Cog study was performed in 2016 and 2021 with 2,042 and 3,575 participants, respectively. The databases contain valuable information from areas in Mexico with the largest migration to the U.S. therefore potentially providing valuable insights into the AD/ADRD markers for Mexican Americans.[6]

Low cognitive reserve and poor gut-biome health have been investigated as risk factors for AD/ADRD. For example, lower levels of education have been associated with and contribute to an increased risk of cognitive decline due to a lack of cognitive reserve.[7] Using ML and the proposed databases, we hope to discover non-biological markers impacting the Mexican population. We propose to study, quantify, and determine correlations between AD/ADRD and social factors. Using ML architectures and Shapley values, we can evaluate novel markers for AD/ADRD[8] that can be used as early predictors of AD/ADRD for the Mexican population. For our analysis, examples of predictor variables are childhood socioeconomic conditions, marital status, fertility, unions or marriages, family migration history, family structure, including number of children and siblings (deceased and living), parents living status and their education, weekly contact with family, money transfer behaviors, and social and religious activities. Additionally, the database also contains variables such as widowhood and its economic consequences, individual income, assets, broad employment, pension history, type of housing, building materials, other indicators of housing quality, property conditions, availability of consumer durables, self-reported health conditions, a cognitive assessment, anthropometric measurements, and blood analysis for a subsample of the data.[9]

Many of the variables are extensions of predictor areas that are being studied as risk / protective factors for AD/ADRD, such as gut biome (diet, hygiene, living conditions, housing quality, obesity),[10,11] lack of neuroplasticity (education, social interaction/isolation, family contact, living alone, hearing loss),[7] contribution of other health issues including metabolic syndrome (obesity, diabetes, dyslipidemia),[12,13,14] lifestyle (smoking, alcohol)[15,16]. To evaluate the impact of the predictor variables on the development of AD/ADRD, we propose as target variables each participant's cognitive impairment assessed through the MMSE and CSI'D scores available on the Mex-Cog database.[5] The data was collected through interviews in person by trained full-time interviewers of the Mexican Statistical Bureau. The data has been digitized and is in English. It is also freely accessible to researchers and suitable for use in ML. The data has no restrictions on sharing due to ownership, data sensitivity, or copyright restrictions.[5]

Most Mexicans migrate to the U.S. between the ages of 24 and 45.[17,18] They bring a lifetime of experiences and environmental exposures that influence health outcomes. Early-life factors such as anthropometric measurements, socioeconomics of childhood, and family dynamics can be crucial in understanding the etiology of AD/ADRD among Mexicans. Moreover, our research and the database can extend to occupational health, exploring the correlation between specific vocations among the participants—such as mining and pottery—and the incidence of AD/ADRD. Using these datasets and ML, we hope to establish such connections and identify at-risk groups that have the potential to inform macro-level health strategies and interventions in the U.S. for the 38 million Mexican Americans that go beyond the five fingers of intervention,[19,20] 1) healthy food, 2) physical activity, 3) mental stimulation, 4) social activates and 5) monitoring of cardiovascular risk factors.

Data Description

1.1 Basic Information

We propose to use the Mexican Health and Aging Study (MHAS) in conjunction with the Mexican Cognitive Aging Ancillary Study (Mex-Cog). The MHAS is a project developed by the National Institute of Statistics, Geography, and Informatics (INEGI) in Mexico and researchers from the Universities of Pennsylvania, Maryland, and Wisconsin in the US[21]. The study aimed to obtain national-level information about the aging process of the population aged 50 and over in Mexico from a broad socioeconomic perspective[22]. MHAS contains six waves of data: a baseline interview in 2001 with follow-ups in 2003, 2012, 2015, 2018, and 2021[6]. The sample was distributed over all 32 states in rural and urban areas, with an over-sample of households in the six states that account for 40% of all migrants to the U.S. The surveys were completed through direct interviews or proxy interviews in cases of illness, hospitalization, or temporary absence of the interviewed. The 2001 and 2003 interviews were conducted using paper and pencil; from 2012 on, the interviews were completed using Computer Assisted Personal Interview (CAPI)[5].

On the other hand, the Mex-Cog 2016 and 2021 studies are an in-depth cognitive assessment applied to a subsample of age 55 and older of MHAS 2015 and MHAS 2018[6,23]. For the 2016 cohort, the sample was distributed in 8 states selected using a stratified sampling procedure. The states were chosen to represent the national population according to socioeconomic conditions (percentage of urban and rural, number of residents who are former migrants to the U.S.) and health exposures (percentage of obesity, diabetes, mining industry, and pottery industry). All MHAS 2015 subjects who were eligible in each state were included[5]. The geographic details of the 2021 cohort are not available yet. For both waves, most of each survey was completed using CAPI; however, paper and pencil were employed for specific cognitive exercises[5].

MHAS until the 2018 cohort and Mex-Cog 2016 were harmonized following the RAND HRS and Harmonized HRS conventions of variable naming and data structure[24]. The harmonized MHAS comprises one file available in STATA (184.7 MB), SAS (481.9 MB), and SPSS (185.5 MB) formats. The Mex-Cog study consists of two files, one for the harmonized 2016 cohort and one not harmonized for the 2021 cohort. The former is available in STATA (1.3 MB), and the latter is available in STATA (3.2 MB), SAS (4.6 MB), and SPSS (1.8 MB). The number of subjects in Mex-Cog 2016 is 2,042, and 3,575 for 2021, with a 24.42% overlap between the two waves. Therefore, there are 4,245 unique participants. Of those subjects, 2,901 were enrolled in MHAS in 2001, 1,262 in 2012, 2 in 2015, and 80 in 2018. Both databases are available at the MHAS website, harmonized databases can be found in this link_1 and non-harmonized in this link_2. The target variable in this database is cognitive impairment which was assess through a modified Mini-Mental State Examination (MMSE) and the Community Screening Instrument for Dementia (CSI'D). Given the lack of harmonization in the Mex-Cog 2021, only the MMSE score was identified as a variable for that cohort. In Mex-Cog 2016, 82,46% of participants had MMSE scores, and all had CSI-D scores. In Mex-Cog 2021, all the participants had MMSE. The total number of observations in the MHAS are 26.839 with 5241 variables. For the Mex-Cog 2016 and 2021, the total number of observations are 2042 and 3575 with 405 and 404 variables respectively. The distribution of the target variables is shown in Figure 1.

1.2 Utility & Rigor

Target variable: Cognitive impairment in Mexican population assessed by the Adapted Mini Mental State Examination (MMSE) and the Community Screening Interview for Dementia (CSI'D).

Cognitive impairment: The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, DSM-5 diagnosis of Mild Neurocognitive Disorder, corresponding to MCI, is made when there is modest impairment in one or more cognitive domains. The individual is still independent in everyday activities, albeit with greater effort. The diagnosis of Major Neurocognitive Disorder, which corresponds to dementia, requires substantial impairment to be present in one or (usually)

more cognitive domains. The impairment must be sufficient to interfere with independence in everyday activities[25].

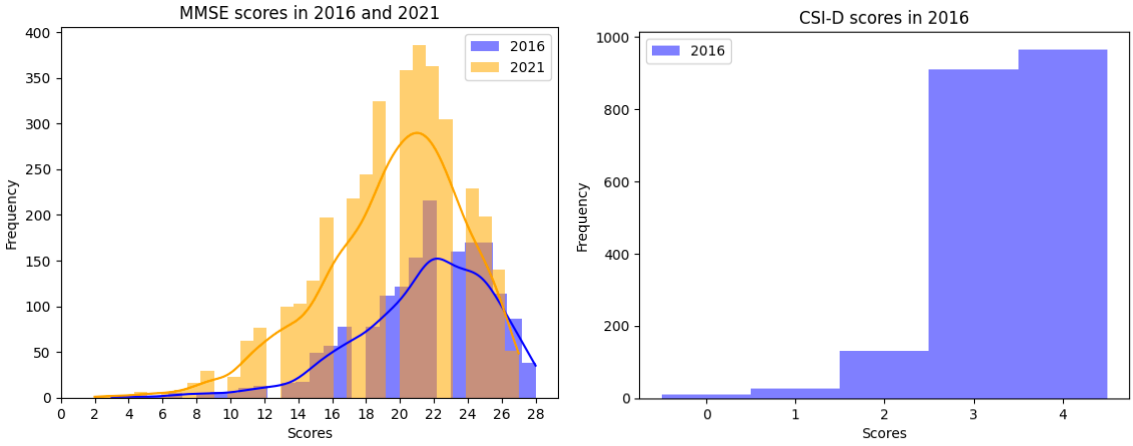

Figure 1. MMSE and CSI-D scores distribution.

Screening tools for providing an overall measure of cognitive impairment:

Mini Mental State Examination (MMSE): The (MMSE) remains as the most widely used instrument of providing a brief screening test that quantitatively assesses the severity of cognitive impairment and documents cognitive changes occurring over time[26]. The MMSE is a screening tool that measures cognitive function by evaluating orientation, attention and calculation, memory, language, and visuospatial skills. To identify cognitive impairment using cut points of (23 or less or 24 or less), the MMSE has a sensitivity of 85 to 92% and a specificity of 85 to 93%[27]. A rough rule of thumb is that patients with mild dementia have scores of 18–26 out of 30, those with moderate dementia scores of 10–18, and those with severe dementia score ≤ of less than 10[28]. This assessment has already been validated in Spanish in Mexico[24], and the Mex-cog has adapted the Mexican-validated MMSE to be more applicable to the general Mexican urban and rural population[5]. The adapted version is a 10-item scale with a total score of 28. The following assessment battery is divided into three parts: the first part intended to detect dementia (verbal learning, verbal fluency, visual scan, backward counting, naming, and delayed verbal recall), the second part intended to evaluate higher cognitive demands for the detection of MCI (immediate and delayed logical memory, list word recognition, copy and recall constructional praxis, and symbol-digit), and the third part intended to measure executive function (abstract reasoning and inhibition control). The final score includes an individual score for each domain and a total composite score for the sum across all domains[5].

A common challenge that cognitive assessments face is that the difficulty experienced by respondents can cause a significant burden by increasing fatigue and irritability and potentially causing less engagement in subsequent therapies.[30] This effect is widespread in populations with low education and literacy levels, revealing biased and highly heterogeneous scores in tasks involving reading, writing, arithmetic, and visuospatial skills.[31,32] The Mex-Cog database takes this issue into consideration by determining which cognitive assessment, long or short, would be better suited to respondents. This means that individuals who obtained ≤ 10 in the adapted MMSE continued with the short assessment, which only included the first part of the assessment battery, and individuals whose scores ranked 11 or more continued with the second and third parts of the assessment battery. The adaptation of the MMSE has shown high internal consistency (Cronbach's alpha = 0.93) with a good item-total correlation for each MMSE domain. Variation of

cognitive abilities by demographics, years of education, and age showed consistent correlations with cognitive decline in the MMSE domains scores.

Community Screening Interview for Dementia (CSI'D): A promising way to screen the cognitive status of participants outweighing educational bias in the final scores can be determined by evaluating the Community Screening Interview for Dementia (CSI'D).[32] The CSI'D includes an interview with the subject (cognitive testing) for non-literate and literate populations and an interview with an informant to assess the subject's performance in everyday living. The original version of the CSI'D has 27 items but was separated into 28 in the Mex-Cog by separating the incontinence question into two questions for exploring urine and bowel control. The Mex-Cog database includes informant's score for cognitive and functional decline in the analysis. Reliability scores for the informant questionnaire showed a high internal consistency (Cronbach's alpha = 0.88), which means that the items measured in the informant's questionnaire consistently assess the subject's performance in everyday activities. The score negatively correlates with the Adapted-MMSE scores in the Mexican population. The CSI'D has previously demonstrated to be adaptable to populations with different socioeconomic backgrounds and levels of education, allowing it to explore Mexican rural and urban contexts[5,32].

Cognitive impairment predictor variables:

Cognitive impairment is often an early sign of Alzheimer's dementia. Longitudinal studies have demonstrated that the rate of cognitive decline can help predict the likelihood of developing dementia in the future[33]. Some people with mild cognitive impairment (MCI) will progress to dementia but others remain stable or recover full function[34]. Among the aged with mild cognitive impairment, almost 1/3 is likely to progress to dementia over 5 years[35]. Understanding the predictor variables that may lead to cognitive impairment is crucial for early prevention of AD/ADRD.

The National Institute on Aging (NIA) Health Disparities Research Framework provides a comprehensive approach to addressing health disparities among older adults. When considering predictors of cognitive impairment and AD/ADRD within this framework, we can find various groups of factors in common in including sociocultural determinants, biological factors, behavioral factors, and environmental influences. The MHAS in conjunction with the Mex-Cog contain the following predictor variables of the groups within this framework:

Group	Predictor Variables
Sociocultural Determinants	childhood socioeconomic conditions, marital status, unions or marriages, family migration history, family structure, including number of children and siblings (deceased and living), parents living status and their education, employment condition, weekly contact with family, transfer behaviors, and social and religious activities (social support), widowhood and its economic consequences, individual income, assets, broad employment, pension history, type of housing, building materials, other indicators of housing quality, property conditions, and availability of consumer durables, access to health Insurance coverage.
Behavioral Factors	smoking cigarettes, alcohol drinking, sleep behaviors
Environmental Factors	living in rural or urban area.
Biological Factors	age, sex, gender, high blood pressure, diabetes, cancer, respiratory disease, incl asthma, heart attack, heart problems, stroke, arthritis, depression, bone fractures, urinary incontinence, hearing loss, COVID-19, stress measures. Self-reported health conditions, a general cognitive assessment, anthropometric measurements, depressive symptoms, and blood biomarkers analysis for a subset of the data are also relevant withing the database.[9]

The factors described above provide valuable insights into the sources of risk and the agency individuals have in controlling their own risk of dementia[7]. In other words, including potential variables to explain the risk for the development of cognitive decline and a later AD/ADRD will be essential to explain the risk and progression of the disease in different contexts. Each group of variables contributes a level of explanation for the scores of cognitive decline and level of functionality obtained in our target variable. As mentioned, all the proposed factors described by the NIA Health Disparities Research Framework allow us to create a model that incorporates at different levels the variables that are more impactful in the overall health and cognitive state of individuals across their lifespan.

We propose that socioeconomic determinants, specifically the associations of education, occupational status (e.g., unemployment), net income, material deprivation, and other material circumstances, will have a significant effect on lifestyle for brain health[36,37]. This may suggest that poorer lifestyle conditions, the experience of discrimination[38], the effect of neighborhood environments, public transportation, and even the quality of public spaces may contribute to more stress levels[39], and more adverse lifestyle decisions and a younger onset of cognitive decline compared to socially advantaged individuals[40]. Interestingly, education seems to be strongly associated with decreasing this risk, as higher education may increase the likelihood of engaging in more cognitively demanding occupations and activities, which would increase individuals' cognitive reserve[41]. This variable has also been associated with the adoption of healthier lifestyles, such as reduced tobacco smoking, reduced alcohol intake, earlier treatment for hypertension and diabetes, a healthy diet, and physical activity promote better cognitive performances and a better overall health compared to more sedentary lifestyles[42]. On the other hand, the effect of social experiences has also been reported, for example, lower levels of social engagement[37], and social isolation are associated with poor cognitive function[43], with depression as a possible mediator[44]. In addition, we also considered that lifestyle social and religious participation will influence health and well-being, as shown by previous studies that have reported the effect of social participation as positive for cognitive health, lowering 30-50% the subsequent dementia risk[45-47].

Additionally, biological factors are a crucial group of variables to be included in the analysis. Blood biomarkers such as vitamin D, cholesterol, glycosylated Hemoglobin, C-reactive protein, Thyroid Stimulating Hormone, support the assessment of the metabolic syndrome, associated with cognitive impairment[48]. Also, pre-existing conditions such obesity, hypertension, diabetes, depression, hearing loss, COVID-19, stress, may be associated with an increased risk of cognitive impairment.

Overall, the proposed databases enable the analysis of various types of risk factors and their possible relationships. Since MHAS is part of the HRS International Family of Studies, and Mex-Cog follows the harmonized protocol for cognitive aging (HCAP), the models designed using these databases can be easily extrapolated to studies following the same protocols. For example, databases such as CHARLS in China, ELSA and ELSA-HCAP in England, LASI and LASI-DAD in India, HAALSI in Africa, and HRS and HCAP in the United States can be concurrently utilized to develop comprehensive models that consider diverse international populations, including those typically underrepresented and at higher risk for AD/ADRD, such as Hispanic and African communities[49].

1.3 Innovation

In the U.S., non-Hispanic Black and Hispanic older adults are disproportionately more likely than White older adults to suffer from AD/ADRD[4]. Data from the Census Bureau and the Chicago Health and Aging Project (CHAP) study indicates 19% of Black and 14% of Hispanic adults age 65 and older have Alzheimer's dementia compared with 10% of White older adults[2]. However, most clinical research is performed with White Americans. Data has suggested that Hispanic/Latino populations represent around 4.4% or less of the total participation in some North American randomized controlled trials (RCT) for Alzheimer's Disease[50]. Comparatively, other

studies have raised the concern of the shrinking minority of enrolled non-Hispanic Black and Hispanic/Latino populations in ARDR clinical trials, with a cumulative 4.5% participation[51]. Recent improvements in recruitment, such as the 2022 Lecanemab in Early Alzheimer's Disease trial, increased Hispanic/Latino enrollment to 12.4%; nevertheless, non-Hispanic Whites were still overrepresented with approximately 76.9% of total enrollment. More concerning was the non-Hispanic Black population, which had a mere 2.5% of total participation in the trial[52]. Underrepresented racial/ethnic communities are often ineligible after initial screenings and, therefore, fail to be enrolled[53]. Various factors contribute to enrollment failure, including socioeconomic status, literacy, acculturation, stigma, discrimination, neighborhood health metrics, language, and beliefs about dementia within these populations[54,55]. This situation raise concerns because the low representation of these communities limits the understanding of the risk factors for developing ADRD. Considering that Latinos are the largest and fastest-growing minority in the United States,[1] numbered 63 million in 2022, with nearly 60% of Mexican origin,[2,3] it is important to develop models tailored to this specific population.

Additionally, utilizing survey-based data to identify and assess risk factors for developing AD/ADRD enables the creation of high-reach tools. These tools would permit the evaluation of a broad segment of the community concerning their potential to develop AD/ADRD. Furthermore, this approach is particularly advantageous as it circumvents the high costs associated with imaging techniques such as MRI and CFS biomarkers analysis. Consequently, it facilitates the evaluation of large societal groups at minimal costs.

Moreover, the amount of socioeconomic and demographic information contained within these databases allows for identifying social and cultural factors that may influence the development and progression of the disease, extending beyond the biological factors that have been more extensively studied. Integrating both types of information—biological and socio-demographic—enhances early detection capabilities by correlating risk factors with early health variations. This means that, for individuals with certain socio-demographic risk factors, minor health changes that might otherwise be considered normal may indicate the disease's onset. Since brain changes can begin years before symptoms arise, detecting these subtle health variations early on, combined with the discovered risk factors, can justify imaging techniques for early detection of AD/ADRD.

Moreover, using this database enhances the understanding of AD/ADRD development within Hispanic (Mexican) populations, a demographic group with a greater prevalence to these conditions and yet underrepresented in United States healthcare research. The utilization of MHAS and Mex-Cog databases offers a distinct advantage over sources like the Health and Retirement Study, as it provides tailored insights into the unique epidemiological patterns of this specific group. This strategic approach enriches the precision of the research and highlights the commitment to inclusive representation in the study of AD/ADRD.

1.4 Sample Characteristics and Representation

The participants on the MHAS were drawn from the National Survey of Employment in Mexico, based on age-eligibility. The next waves followed up with the original sample, but also including new spouses and next-of-kin responses when participants passed away. There were two refreshment samples that were drawn from the National Occupation and Employment Survey in 2012 and 2018. Even though MHAS sample was collected over the 32 states in Mexico, Mex-Cog 2016 sample was distributed in 8 states selected using a stratified sampling procedure. The states were chosen to represent the national population according to socioeconomic conditions and health exposures. Sampling details of the 2021 Mex-Cog aren't available yet. Regarding some important characteristics of the sample, figure 2 depict that the average age on inclusion at MHAS was 57.16 ± 6.64 years, for the Mex-Cog study it was 71.19 ± 8.55 years.

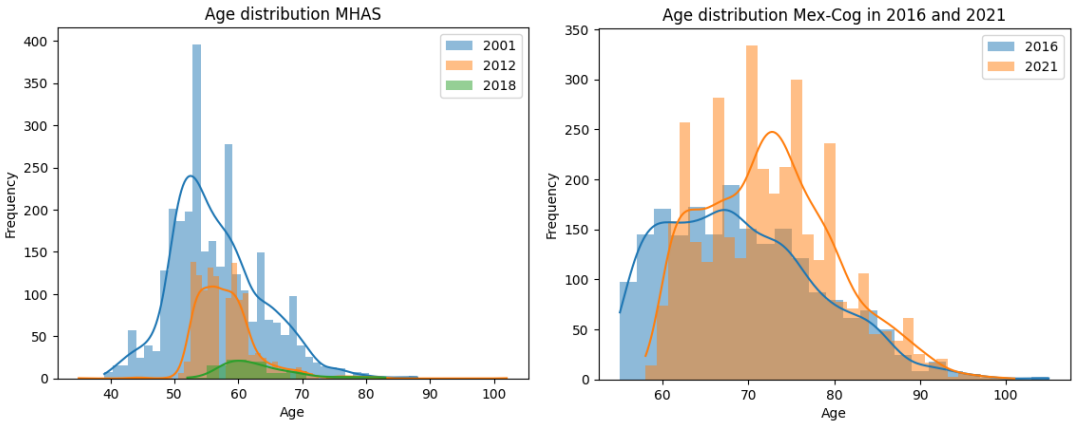

Figure 2. Participants' age distributions.

Concerning sex, 57.1% of the sample is comprised by women. Regarding cognitive reserve, some indicators can be education level, literacy and numeracy, and speaking other languages. The average years of education for participants was 5.22 years with a standard deviation of 4.45 years. This result shows that many of the participants didn't finish primary or secondary level of education. However, most participants reported to be literate (83.6%), able to count and use numbers (94.1%), with around 10% able to speak an indigenous language.
Finally, evaluated at the last wave, around half of the participants (50.7%) lived in urban households, while 22% lived either in semi-urban or semi-rural households, and 21% in rural. More than half of the participants (52.5%) lived in coupled households. As mentioned before, this database enables the analysis of the largest and fastest-growing minority in the US, a demographic group with a greater predisposition to develop AD/ADRD and yet underrepresented in United States healthcare research.

1.5 Usability

The MHAS and Mex-Cog databases have already been de-identified. All data and documentation are stored at the MHAS website[56] and are of public use if researchers register on the website and agree to its terms of use. The terms are the following: will not attempt to identify participants, will not distribute MHAS username and password, will not transfer MHAS public release data to third parties, except to staff or students for whom said user is directly responsible, and will adequately credit MHAS for the use of any data and will include the link to the public data in any publication[23].

As mentioned before the MHAS data was harmonized following the RAND HRS and Harmonized HRS conventions of variable naming and data structure until the 2018 cohort[57]; therefore, information about the 2021 follow-up isn't ready to be used. Furthermore, since Mex-Cog 2021 hasn't been harmonized, only the MMSE target variable is available for that cohort; the CSI'D score wasn't identified within the raw data variables.

2 Team Introduction

IGC Pharma team is comprise of employees of the company encompassing professionals with specialized expertise in Alzheimer's disease, and experts in Artificial Intelligence and Machine Learning (AI/ML):

Ram Mukunda (CEO – Team Lead): Over the past nine years, he has performed research and lead teams within the medical and pharmaceutical industry. His desire is to develop low-cost medications for AD and develop early markers of AD. Currently, he leads a multisite, placebo-controlled Phase 2b clinical trial on agitation in AD and the AI/ML team. BS Mathematics (1979), BS/MS Electrical Engineering (1979,1982), University of Maryland.
Pablo Arbeláez (AI/ML Expert): Extensive experience using AI/ML in medicine, biology, and computer vision. Recognized with the Artificial Intelligence 2000 Most influential Scholar Award and awards that place him among the 100 most influential researchers in this discipline in the last decade. PhD Applied Mathematics, Universite Paris-Dauphine (2005), AI researcher University of California (Berkeley 2007-2014), Professor of Biomedical Engineering, University of Los Andes, Colombia.
Juan Manuel Orjuela (Medical Director): He is a neuropsychiatrist responsible for the clinical aspects of our AD-related trial. He works on the clinical trial team, as he has a vast knowledge of neuropsychiatric disorders.
Evelyn Gutiérrez (Scientific Manager): She is a Chemical Engineer and is currently pursuing a Master of science (MSc) in clinical Epidemiology. Training in GCP, Clinical Monitoring, and statistical analysis in STATA. She has four years of experience in Clinical Research focused on AD and is leading the IGC Pharma Clinical trial team.
***Paola Ruíz (Data Scientist*)**: She is a Biomedical Engineer who focuses on using AI/ML to address medical challenges. She has more than four years of experience using deep learning (DL) for drug discovery, high-definition image analysis, and understanding of surgical scenes. She has expanded her research to enhancing AD/ADRD diagnosis by means of ML tools.
Daniel Crovo (AI Research Engineer): He is an Electronics Engineer currently pursuing a master's degree in AI/ML. Experienced in developing deep learning models for medical image analysis. His master's thesis explores intracranial EEG signal analysis. He is committed to applying AI in the medical field research, particularly for early Alzheimer's detection.
Nestor González (AI Research Engineer): He is a computer sciences engineer, with experience applying AI to the medical field. He has used ML to identify lesions associated with drug-resistant epilepsy through MRI scans. Additionally, he has applied DL to predict real time changes in trust dynamics between humans and technology, based on physiological data.
Juanita Arbeláez (Medical Doctor): She is a medical doctor with a master's in Epidemiology and currently pursuing a master's in bioethics. Her specific interests in neurology, particularly in neurodegenerative diseases, guide her impactful research on AD/ADRD and related neuropsychiatric symptoms. She integrates her skills in scientific writing and her expertise in creating and managing databases, alongside statistical analysis and interpretation of results.
Margarita Venegas (Clinical Psychologist): She is a clinical and health psychologist with a comprehensive understanding of clinical assessment, and quantitative and qualitative research. She has experience in the assessment processes to apply and evaluate numerous psychological scales. She also reviews data related to the scales.
Alejandra Tangarife (Neuroscientist): She is a neuroscientist specialized in developmental psychology, social neuroscience, and cognitive neuroscience. She focuses on processes related to working memory, language, and complex behaviors associated with socialization and social cognition. Her knowledge of psychological scales will allow her to integrate biological and sociocultural factors to increase awareness concerning the structural difficulties surrounding the underrepresentation of diverse ethnic/racial communities in clinical research practice.

3 References

1. González Burchard, E., Borrell, L. N., Choudhry, S., Naqvi, M., Tsai, H. J., Rodriguez-Santana, J. R., Chapela, R., Rogers, S. D., Mei, R., Rodriguez-Cintron, W., Arena, J. F., Kittles, R., Perez-Stable, E. J., Ziv, E., & Risch, N. (2005). Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. American journal of public health, 95(12), 2161–2168. https://doi.org/10.2105/AJPH.2005.068668
2. National Population by Characteristics: 2020-2023. (2023, December 28). United States Census Bureau. https://www.census.gov/data/datasets/time-series/demo/popest/2020s-national-detail.html
3. Krogstad, J.M., Passel, J.S., Moslimani, M., & Noe-Bustamante, L. (2023, September 22). Key facts about U.S. Latinos for National Hispanic Heritage Month. Pew Research Center. https://www.pewresearch.org/short-reads/2023/09/22/key-facts-about-us-latinos-for-national-hispanic-heritage-month
4. 2023 Alzheimer's disease facts and figures. (2023). Alzheimer's & dementia : the journal of the Alzheimer's Association, 19(4), 1598–1695. https://doi.org/10.1002/alz.13016
5. Mejia-Arango, S., Nevarez, R., Michaels-Obregon, A., Trejo-Valdivia, B., Mendoza-Alvarado, L. R., Sosa-Ortiz, A. L., Martinez-Ruiz, A., & Wong, R. (2020). The Mexican Cognitive Aging Ancillary Study (Mex-Cog): Study Design and Methods. Archives of gerontology and geriatrics, 91, 104210. Advance online publication. https://doi.org/10.1016/j.archger.2020.104210
6. Wong, R., Michaels-Obregon, A., & Palloni, A. (2017). Cohort Profile: The Mexican Health and Aging Study (MHAS). International journal of epidemiology, 46(2), e2. https://doi.org/10.1093/ije/dyu263
7. Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., Cooper, C., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Larson, E. B., Ritchie, K., Rockwood, K., Sampson, E. L., Samus, Q., … Mukadam, N. (2017). Dementia prevention, intervention, and care. Lancet (London, England), 390(10113), 2673–2734. https://doi-org/10.1016/S0140-6736(17)31363-6
8. Sun, T., Chen, H., Qiu, Y., Zhao, C. (2023). Efficient Shapley Values Calculation for Transformer Explainability. In: Lu, H., Blumenstein, M., Cho, SB., Liu, CL., Yagi, Y., Kamiya, T. (eds) Pattern Recognition. ACPR 2023. Lecture Notes in Computer Science, vol 14406. Springer, Cham. https://doi.org/10.1007/978-3-031-47634-1_5
9. Hasan, N., & Yang, H. (2019). Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ*, *7*, e7502. https://doi.org/10.7717/peerj.7502
10. Ferreiro, A. L., Choi, J., Ryou, J., Newcomer, E. P., Thompson, R., Bollinger, R. M., Hall-Moore, C., Ndao, I. M., Sax, L., Benzinger, T. L. S., Stark, S. L., Holtzman, D. M., Fagan, A. M., Schindler, S. E., Cruchaga, C., Butt, O. H., Morris, J. C., Tarr, P. I., Ances, B. M., & Dantas, G. (2023). Gut microbiome composition may be an indicator of preclinical Alzheimer's disease. *Science translational medicine*, *15*(700), eabo2984. https://doi.org/10.1126/scitranslmed.abo2984
11. Gilsanz, P., Mayeda, E. R., Glymour, M. M., Quesenberry, C. P., Mungas, D. M., DeCarli, C., Dean, A., & Whitmer, R. A. (2017). Female sex, early-onset hypertension, and risk of dementia. Neurology, 89(18), 1886–1893. https://doi-org/10.1212/WNL.0000000000004602
12. Gottesman, R. F., Albert, M. S., Alonso, A., Coker, L. H., Coresh, J., Davis, S. M., Deal, J. A., McKhann, G. M., Mosley, T. H., Sharrett, A. R., Schneider, A. L. C.,

Windham, B. G., Wruck, L. M., & Knopman, D. S. (2017). Associations Between Midlife Vascular Risk Factors and 25-Year Incident Dementia in the Atherosclerosis Risk in Communities (ARIC) Cohort. JAMA neurology, 74(10), 1246–1254. https://doi-org/10.1001/jamaneurol.2017.1658

13. Katon, W., Pedersen, H. S., Ribe, A. R., Fenger-Grøn, M., Davydow, D., Waldorff, F. B., & Vestergaard, M. (2015). Effect of depression and diabetes mellitus on the risk for dementia: a national population-based cohort study. JAMA psychiatry, 72(6), 612–619. https://doi-org/10.1001/jamapsychiatry.2015.0082
14. Alonso, A., Mosley, T. H., Jr, Gottesman, R. F., Catellier, D., Sharrett, A. R., & Coresh, J. (2009). Risk of dementia hospitalization associated with cardiovascular risk factors in midlife and older age: the Atherosclerosis Risk in Communities (ARIC) study. *Journal of neurology, neurosurgery, and psychiatry*, *80*(11), 1194–1201. https://doi-org/10.1136/jnnp.2009.176818
15. Chen, R., Wilson, K., Chen, Y., Zhang, D., Qin, X., He, M., Hu, Z., Ma, Y., & Copeland, J. R. (2013). Association between environmental tobacco smoke exposure and dementia syndromes. *Occupational and environmental medicine*, *70*(1), 63–69. https://doi-org.ezproxy.uniandes.edu.co/10.1136/oemed-2012-100785
16. Rosenblum, M. R., Kandel, W. A., Ribando Seelke, C., Wasem, R. E. (2012). Mexican Migration to the United States: Policy and Trends. Washington, D.C.: Congressional Research Service.
17. Publisher U.S. Census Bureau. (2023, July 19). Population estimates: Estimates by age group, sex, race, and Hispanic origin. United States Census Bureau. https://catalog.data.gov/dataset/population-estimates-estimates-by-age-group-sex-race-and-hispanic-origin
18. Ngandu, T., Lehtisalo, J., Solomon, A., Levälahti, E., Ahtiluoto, S., Antikainen, R., Bäckman, L., Hänninen, T., Jula, A., Laatikainen, T., Lindström, J., Mangialasche, F., Paajanen, T., Pajala, S., Peltonen, M., Rauramaa, R., Stigsdotter-Neely, A., Strandberg, T., Tuomilehto, J., … Kivipelto, M. (2015). A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (finger): A randomised controlled trial. *The Lancet*, *385*(9984), 2255–2263. https://doi.org/10.1016/s0140-6736(15)60461-5
19. Rosenberg, A., Mangialasche, F., Ngandu, T., Solomon, A., & Kivipelto, M. (2020). Multidomain Interventions to Prevent Cognitive Impairment, Alzheimer's Disease, and Dementia: From FINGER to World-Wide FINGERS. The journal of prevention of Alzheimer's disease, 7(1), 29–36. https://doi.org/10.14283/jpad.2019.41
20. Estudio Nacional de Salud y Envejecimiento en México 2001 (ENASEM) - Documento Metodológico Reporte de Proyecto. (2004, June 11). https://www.mhasweb.org/resources/DOCUMENTS/2001/Methodological_Document_2001.pdf
21. *Study on Cognitive Aging Linked to MHAS: Methodological Document, Version 2*. (2020). Mex-Cog. https://www.mhasweb.org/resources/DOCUMENTS/2015/Mex-Cog/Methodological_Document_Mex_Cog_2016.pdf
22. *Study Description*. (n.d.). Mexican Health & Aging Study. *https://www.mhasweb.org/Home/StudyDescription.aspx*
23. Michaels-Obregon, Alejandra, Drystan Phillips, Jenny Wilkens, Rebeca Wong, and Jinkook Lee. “Harmonized MHAS, Version C.2.” Gateway to Global Aging Data, 2023. https://doi.org/10.34729/06K4-ZG60
24. Reyes de Beaman, S., Beaman P.E., Garcia-Peña, D., Villa, M.A., Heres, J., Córdova, A., & Jagger, C. (2004). Validation of a Modified Version of the Mini-Mental State

Examination (MMSE) in Spanish, Aging. *Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development, 11*(1), pp.1-11, https://doi.org/10.1076/anec.11.1.1.29366

25. Hugo, J., & Ganguli, M. (2014). Dementia and Cognitive Impairment: Epidemiology, Diagnosis, and Treatment. Clinics in Geriatric Medicine, 30(3), 421. https://doi.org/10.1016/j.cger.2014.04.001
26. Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: a comprehensive review. Journal of the American Geriatrics Society, 40(9), 922–935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x
27. Patnode CD, Perdue LA, Rossom RC, Rushkin MC, Redmond N, Thomas RG, Lin JS: Screening for Cognitive Impairment in Older Adults: An Evidence Update for the U.S. Preventive Services Task Force. In. Rockville (MD); 2020.
28. Feldman, H. H., Jacova, C., Robillard, A., Garcia, A., Chow, T., Borrie, M., Schipper, H. M., Blair, M., Kertesz, A., & Chertkow, H. (2008). Diagnosis and treatment of dementia: 2. Diagnosis. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 178(7), 825–836. https://doi.org/10.1503/cmaj.070798
29. Tomaszczyk, J., Sharma, B., Chan, A. A., Colella, B., Mok, J., Beaton, D., Christensen, B., & Green, R. (2018). Measuring cognitive assessment and intervention burden in patients with acquired brain injured: Development of the "How Much is Too Much" questionnaire. Journal of rehabilitation medicine, 50(6), 519–526. https://doi.org/10.2340/16501977-2344
30. Pellicer-Espinosa, I., & Díaz-Orueta, U. (2022). Cognitive Screening Instruments for Older Adults with Low Educational and Literacy Levels: A Systematic Review. Journal of applied gerontology: the official journal of the Southern Gerontological Society, 41(4), 1222–1231. https://doi.org/10.1177/07334648211056230
31. Brucki, S. M. D., & Nitrini, R. (2010). Mini-Mental State Examination among lower educational levels and illiterates: Transcultural evaluation. Dementia & neuropsychologia, 4(2), 120–125. https://doi.org/10.1590/S1980-57642010DN40200008
32. Hall, K.S., Gao, S., Emsley, C.L., Ogunniyi, A.O., Morgan, O., & Hendrie, H.C. (2000). Community screening interview for dementia (CSI 'D'); performance in five disparate study sites. International Journal of Geriatric Psychiatry, 15(6), 521–531. doi:10.1002/1099-1166(200006)15:6
33. Lim, Y. Y., Kong, J., Maruff, P., Jaeger, J., Huang, E., & Ratti, E. (2022). Longitudinal Cognitive Decline in Patients With Mild Cognitive Impairment or Dementia Due to Alzheimer's Disease. The journal of prevention of Alzheimer's disease, 9(1), 178–183. https://doi.org/10.14283/jpad.2021.64
34. Arevalo-Rodriguez, I., Smailagic, N., Roqué-Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O. L., Cosp, X. B., & Cullum, S. (2021). Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). The Cochrane Database of Systematic Reviews, 2021(7). https://doi.org/10.1002/14651858.CD010783.pub3
35. Su, Y., Dong, J., Sun, J., Zhang, Y., Ma, S., Li, M., Zhang, A., Cheng, B., Cai, S., Bao, Q., Wang, S., & Zhu, P. (2021). Cognitive function assessed by Mini-mental state

examination and risk of all-cause mortality: a community-based prospective cohort study. BMC geriatrics, 21(1), 524. https://doi.org/10.1186/s12877-021-02471-9

36. Röhr, S., Pabst, A., Baber, R., Engel, C., Glaesmer, H., Hinz, A., Schroeter, M. L., Witte, A. V., Zeynalova, S., Villringer, A., Löffler, M., & Riedel-Heller, S. G. (2022). Social determinants and lifestyle factors for brain health: implications for risk reduction of cognitive decline and dementia. *Scientific reports*, *12*(1), 12965. https://doi.org/10.1038/s41598-022-16771-6
37. Majoka, M. A., & Schimming, C. (2021). Effect of Social Determinants of Health on Cognition and Risk of Alzheimer Disease and Related Dementias. Clinical therapeutics, 43(6), 922–929. https://doi.org/10.1016/j.clinthera.2021.05.005
38. Zuelsdorff, M., Okonkwo, O. C., Norton, D., Barnes, L. L., Graham, K. L., Clark, L. R., Wyman, M. F., Benton, S. F., Gee, A., Lambrou, N., Johnson, S. C., & Gleason, C. E. (2020). Stressful Life Events and Racial Disparities in Cognition Among Middle-Aged and Older Adults. Journal of Alzheimer's disease : JAD, 73(2), 671–682. https://doi.org/10.3233/JAD-190439
39. Aggarwal, N. T., Wilson, R. S., Beck, T. L., Rajan, K. B., Mendes de Leon, C. F., Evans, D. A., & Everson-Rose, S. A. (2014). Perceived stress and change in cognitive function among adults 65 years and older. Psychosomatic medicine, 76(1), 80–85. https://doi.org/10.1097/PSY.0000000000000016
40. Hale, J. M., Schneider, D. C., Mehta, N. K., & Myrskylä, M. (2020). Cognitive impairment in the U.S.: Lifetime risk, age at onset, and years impaired. SSM - population health, 11, 100577. https://doi.org/10.1016/j.ssmph.2020.100577
41. Sonnen, J., Strauss, M., Schneider, J. A., Bennett, D. A., & Mungas, D. (2011). Cognitive activities during adulthood are more important than education in building reserve. Journal of the International Neuropsychological Society : JINS, 17(4), 615–624. https://doi.org/10.1017/S1355617711000014
42. Sharp, E. S., & Gatz, M. (2011). Relationship between education and dementia: an updated systematic review. Alzheimer disease and associated disorders, 25(4), 289–304. https://doi.org/10.1097/WAD.0b013e318211c83c
43. Peterson, R. L., Fain, M. J., A Butler, E., Ehiri, J. E., & Carvajal, S. C. (2020). The role of social and behavioral risk factors in explaining racial disparities in age-related cognitive impairment: a structured narrative review. Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition, 27(2), 173–196. https://doi.org/10.1080/13825585.2019.1598539
44. Cardona, M., & Andrés, P. (2023). Are social isolation and loneliness associated with cognitive decline in ageing?. Frontiers in aging neuroscience, 15, 1075563. https://doi.org/10.3389/fnagi.2023.1075563
45. Kuiper, J. S., Zuidersma, M., Zuidema, S. U., Burgerhof, J. G., Stolk, R. P., Oude Voshaar, R. C., & Smidt, N. (2016). Social relationships and cognitive decline: a systematic review and meta-analysis of longitudinal cohort studies. International journal of epidemiology, 45(4), 1169–1206. https://doi.org/10.1093/ije/dyw089
46. Sommerlad, A., Kivimäki, M., Larson, E. B., Röhr, S., Shirai, K., Singh-Manoux, A., & Livingston, G. (2023). Social participation and risk of developing dementia. *Nature aging*, *3*(5), 532–545. https://doi.org/10.1038/s43587-023-00387-0
47. Huang, Z., Guo, Y., Ruan, Y., Sun, S., Lin, T., Ye, J., Li, J., He, L., Wang, S., Shi, Y., & Wu, F. (2020). Associations of Lifestyle Factors With Cognition in Community-

Dwelling Adults Aged 50 and Older: A Longitudinal Cohort Study. Frontiers in aging neuroscience, 12, 601487. https://doi.org/10.3389/fnagi.2020.601487

48. Wang, X., Ji, L., Tang, Z., Ding, G., Chen, X., Lv, J., Chen, Y., & Li, D. (2021). The association of metabolic syndrome and cognitive impairment in Jidong of China: a cross-sectional study. BMC endocrine disorders, 21(1), 40. https://doi.org/10.1186/s12902-021-00705-w
49. *Home* (n.d.). Mexican Health & Aging Study. https://www.mhasweb.org
50. Aranda, M. P., Marquez, D. X., Gallagher-Thompson, D., Pérez, A., Rojas, J. C., Hill, C. V., Reyes, Y., Dilworth-Anderson, P., & Portacolone, E. (2023). A call to address structural barriers to Hispanic/Latino representation in clinical trials on Alzheimer's disease and related dementias: A micro-meso-macro perspective. Alzheimer's & dementia (New York, N. Y.), 9(2), e12389. https://doi.org/10.1002/trc2.12389
51. Canevelli, Marco; Bruno, Giuseppe; Grande, Giulia; Quarata, Federica; Raganato, Riccardo; Remiddi, Francesca; Valletta, Martina; Zaccaria, Valerio; Vanacore, Nicola; Cesari, Matteo (2019). Race reporting and disparities in clinical trials on Alzheimer's disease: a systematic review. Neuroscience & Biobehavioral Reviews, (), S0149763418306754–. doi:10.1016/j.neubiorev.2019.03.020
52. van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer, L. D., & Iwatsubo, T. (2023). Lecanemab in Early Alzheimer's Disease. The New England journal of medicine, 388(1), 9–21. https://doi.org/10.1056/NEJMoa2212948
53. Raman, R., Quiroz, Y. T., Langford, O., Choi, J., Ritchie, M., Baumgartner, M., Rentz, D., Aggarwal, N. T., Aisen, P., Sperling, R., & Grill, J. D. (2021). Disparities by Race and Ethnicity Among Adults Recruited for a Preclinical Alzheimer Disease Trial. JAMA network open, 4(7), e2114364. https://doi.org/10.1001/jamanetworkopen.2021.14364
54. Gill, T. M., Zang, E. X., Murphy, T. E., Leo-Summers, L., Gahbauer, E. A., Festa, N., Falvey, J. R., & Han, L. (2021). Association Between Neighborhood Disadvantage and Functional Well-being in Community-Living Older Persons. JAMA internal medicine, 181(10), 1297–1304. https://doi.org/10.1001/jamainternmed.2021.4260
55. Zuelsdorff, M., Larson, J. L., Hunt, J. F. V., Kim, A. J., Koscik, R. L., Buckingham, W. R., Gleason, C. E., Johnson, S. C., Asthana, S., Rissman, R. A., Bendlin, B. B., & Kind, A. J. H. (2020). The Area Deprivation Index: A novel tool for harmonizable risk assessment in Alzheimer's disease research. Alzheimer's & dementia (New York, N. Y.), 6(1), e12039. https://doi.org/10.1002/trc2.12039
56. *Data Products* (n.d.). Mexican Health & Aging Study. https://www.mhasweb.org/DataProducts/Home.aspx
57. *HRS International Family of Studies and the Harmonized Cognitive Assessment Protocol*. (n.d.). National Institute of Aging. https://www.nia.nih.gov/research/dbsr/global-aging/hrs-international-family-studies-and-harmonized-cognitive-assessment-protocol