Exploring Acoustic Biomarkers of Alzheimer's Disease and Related Dementias on DementiaBank

Executive summary

Alzheimer's Disease and Related Dementias (AD/ADRD) are a global public health concern, impacting over 55 million individuals worldwide, with over 6 million in the United States alone [1]. Early detection is pivotal for effective intervention [2], yet remains challenging, particularly in underserved communities where the disease burden is high and diagnostic resources are scarce [3]. **Our study seeks to address this disparity by investigating digital voice biomarkers as an efficient, non-invasive, and widely accessible early detection tool for AD/ADRD.** Since speech and language changes are among the earliest signs of cognitive decline in these diseases [4], digital voice biomarkers offer a valuable resource for early detection, especially in underrepresented populations.

We utilize the DementiaBank dataset [5], an extensive collection of speech samples from healthy individuals and those with AD/ADRD, predominantly sourced from structured clinical interviews, either over telephone or in-person. DementiaBank adheres to strict ethical standards, including participant consent and privacy, with access restricted to consortium members [5]. Researchers worldwide can get access by submitting the application.

To address DementiaBank's variability and the issue of complexity and poor organization, for example, different corpus data varies in size, scope, and type of information, and the provided information is scattered in various locations (different folders and types of files), we have undertaken a comprehensive curation of the DementiaBank data by **structuring and unifying the diverse datasets** to establish consistency across the board, and thoroughly **organizing the demographic, geographic, and typical clinical data** it contained. These efforts have significantly improved the dataset's usability, ensuring that it offers a more coherent and comprehensive resource.

The target variable is **the label of the individual's cognitive status**, distinctly categorizing participants as either healthy or afflicted with AD/ADRD (we also detailed their AD/ADRD status by grouping them into AD, Probably AD, MCI (mild cognitive impairment) and PPA (primary progressive aphasia)). This critical classification is determined through the interview of the participants or their proxy, or the cognitive tests.

Central to our desired predictive model are the **acoustic features** extracted from the participants' audios [6, 7]. Those features, such as pitch, formants, and speech rate [8], provide a detailed acoustic profile, capturing subtle variations in speech that are indicative of early cognitive changes associated with AD/ADRD and will serve as the predictor variables for this challenge.

Our designed methodology will capitalize on the established field of voice analysis for ADRD detection, **offering a novel dataset that supports the creation of language-agnostic machine learning (ML) models** that are trained and fine-tuned to recognize specific speech patterns to identify early signs of AD/ADRD. By incorporating a rich array of demographic variables, including age, gender, race, and education, along with ADRD subtypes and cognitive assessment data, we aim to **refine the accuracy and fairness of ADRD detection and enhance inclusivity and adaptability across diverse populations**. Our approach can be easily implemented in various settings, including telehealth platforms, **making it more accessible to a wider population and under-resourced communities**. Also, it's attuned to technical nuances like sampling rates and resolution, **enhancing the reliability and precision** of our analysis and **the continuous evolution** of the models as new data is integrated.

Data Description

Basic Information

Our curation for this challenge is based on DementiaBank dataset, an online repository hosting data from individuals who have undergone cognitive and neurological assessments to determine the presence of AD/ADRD. This resource encompasses a variety of corpora and includes both diagnosed cases and cognitively healthy controls.

Our team conducted an extensive evaluation of each corpus to ascertain its relevance to research objectives of this challenge. The selection criteria were predicated on the depth of information provided, including the size of the participant cohort, the diversity of assessments administered, and the completeness of demographic data. Some excluded examples are that have limited pool size which is single digit and lack basic demographic information such as sex or age.

Table 1 provides a consolidated overview of the corpora included in this study. Each corpus records participants' voice (during the interview or cognitive tests) in MP3 format, which is complemented with a transcript file archived in the CHAT (.CHA) format. (We only consider audio and not consider video here, and the participants will be excluded if they don't have valid audio recordings.) Some corpora store their demographic and cognitive tests results in a separate Excel file or Word file, WLS [9] even provides extra recordings and information regarding their tests in the form of zip packages that total more than 10GB.

Out of 2,476 participants across the corpora, 2086 are selected as usable data after excluding those with incomplete interview or assessments. The final cohort comprised 1166 control subjects, 441 with probable AD, 168 with AD, 270 with MCI, and 41 with PPA.

Data for this study were sourced from the DementiaBank (https://dementia.talkbank.org/), which features a range of corpora with varied experimental designs. These include language-based tasks such as verbal fluency [10], sentence construction [11], picture descriptions (e.g., cookie theft picture) [12], and story recall [13], cognitive tests such as word recall tests [14], MoCA [15], MMSE [16], and digit ordering [17]. Additional data comes from conversations and interviews.

The use of DementiaBank data is governed by the Creative Commons CC BY-NC-SA 3.0 copyright license [18, 19]. No associated costs are present. Access is password-protected over the web to both transcripts and media.

Utility & Rigor

The target variable is **the label of the individual's cognitive status** — healthy and AD/ADRD (including groups of AD, Probably AD, MCI and PPA) — and is determined through the language-based tasks, cognitive tests, conversational interactions and interviews (mentioned above) and interviews with the participants or their proxies. Also, we have stated that the final cohort comprised 1166 control subjects and 920 AD/ADRD participants (including 441 with probable AD, 168 with AD, 270 with MCI, and 41 with PPA).

The dataset's reliability and validity in indicating AD/ADRD are strengthened by its large sample size and the inclusion of a substantial control group. By integrating quantitative task results with qualitative interview data, the data collection methodology in DementiaBank ensures a comprehensive assessment of cognitive status, capturing the nuanced spectrum of AD/ADRD. We also curated the datasets to exclude inapplicable samples, such as NA values on structured data or samples without audio files, and excluded small and insufficiently informative corpora, allowing for a reliable and valid representation of target variable.

Table 1 Overview of the Samples

Corpus	Subjects number					Gender		Age			Specific Variables Included	Location of the corpus	Audio File (.MP3) Information		
	AD/ ADRD				Control	Male	Female	Min	Max	Mean			Bit Depth (Kbps)	Sampling Rate (khz)	Audio Size
	AD	MCI	Probable AD	PPA											
Baycrest	3	7	0	10	20	23	17	44	90	69	Education, MoCA, Handedness	Toronto, ON, Canada	192	44.1	<73MB
Delaware	51	0	0	0	31	27	54	61	91	71	Education, Race, MoCA, Occupation	Newark, Delaware, USA	128	32	<13MB
Lu	24	2	0	0	26	20	32	47	99	79	NA	Greenwich, Connecticut, USA	128	44.1	<5MB
Pitt	9	42	255	0	243	206	343	46	90	68	MMSE	Pittsburgh, Pennsylvania, USA	128	44.1	<86MB
VAS	29	35	0	0	34	47	53	65	94	74	MoCA,gai, gds	Lebanon, New Hampshire, USA and Chapel Hill, North Carolina, USA	128	48	<3MB
Hopkins	0	0	0	27	0	11	16	51	87	61	Education, Race, Handedness	Baltimore, Maryland, USA	258	44.1	<10MB
Ivanova	73	90	0	0	197	112	248	50	96	77	MMSE, Education	Salamanca, Spain	128	44.1	<3MB
Ye	0	43	0	0	0	24	19	46	80	62	Education Race	Beijing, China	64	22.05	<3MB
WLS	30	0	186	0	615	382	448	56	87	79	Sex, Entry Age, Education, Diagnosis, Stroke	Wisconsin, USA	128	44.1	<3MB

We would like to utilize **eGeMAPS (extended Geneva Minimalistic Acoustic Parameter Set) V02 [20] as the predictor variables**. This is a standard set of acoustic features designed for voice research, particularly in paralinguistic and health-related studies.

The typical predictor variables examples included in the eGeMAPS v02 are frequency features (e.g., fundamental frequency), energy and amplitude features (e.g., the standard deviation of amplitude), spectral features (e.g., harmonics-to-noise ratio), temporal features (e.g., speech rate and articulation rate), dynamic features (e.g., the temporal changes of frequency and amplitude), voice quality features (e.g., jitter and shimmer), and paralinguistic features (e.g., emotion score)

Theoretically, AD/ADRD are known to impact cognitive functions that are crucial for speech production and language processing. eGeMAPS V02 encompasses a range of acoustic features that theoretically correspond to the physiological and cognitive processes affected by AD/ADRD. For instance, changes in speech rate or pause patterns may reflect slowed cognitive processing, while alterations in pitch variability might indicate reduced emotional expressiveness. Specifically, empirical research has shown that patients with AD/ADRD exhibit changes in acoustic properties, which can be captured by eGeMAPS V02 and linked to the characteristic of AD/ADRD.

The study's reproducibility and robustness can be enhanced through cross-validation within the dataset, ensuring the reliability of ML models. Additional validation can involve comparing results with other acoustic feature sets, like Emobase [6] and ComParE [6], and applying various ML models (e.g., Random Forest, Support Vector Machine, and Neural Networks).

Our curation is under the consideration of various demographic, geographic, and clinical data, which facilitates validation across different demographics and linguistic backgrounds, and will be easily generalized to other voice datasets, especially the communities from different geographic regions, languages, or demographic groups. Moreover, our curated DementiaBank data, providing behaviorally anchored insights, complements traditional clinical assessments, fostering a more comprehensive understanding of AD/ADRD, potentially improving detection precision.

Innovation

Our curation focuses on structuring and unifying the varied corpus in DementiaBank while organizing demographic, geographic, and clinical data to enhance its usability. This curation transforms the dataset into a more coherent and comprehensive resource, facilitates more accurate and earlier AD/ADRD detection and streamlines the use of ML models, leading to more cost-effective data analysis, particularly benefiting under-resourced communities.

Utilizing eGeMAPS V02 for acoustic feature analysis, our study aims to enhance the detection of early-stage ADRD by focusing on speech changes, a key early symptom. Recognizing that voice analysis for ADRD detection is established, **our dataset's novelty lies in facilitating the development of "language-agnostic" models by concentrating on universal acoustic and speech features, irrespective of the spoken language. It encompasses detailed data on ADRD subtypes and clinical information like cognitive assessments, enabling the study of comorbidities. And it includes detailed demographic data on age and gender, as well as race and education levels, to assess and ensure the fairness of AI-driven ADRD detection models. The dataset also accounts for technical variances, such as sampling rates and resolution, to mitigate potential confounding effects in the data analysis.**

The design of utilization of voice data and integration of ML models, offers lower costs in data analysis, and enhances accessibility to relevant information (such as audio collection) and remote applications, which democratizes access to crucial data and broadens participation, especially from underrepresented or remotely located populations. Our novel design on the curated dataset demonstrates significant potential in improving early AD/ADRD prediction and allows for the exploration of novel insights into the progression and characteristics of AD/ADRD, potentially leading to breakthroughs in understanding and treating the disease.

The AZTIAHO Database [21], developed by Lopez-De-Ipina, K. et al., serves as a relevant comparison to our work with DementiaBank. It is a multicultural and multilingual database containing 20 hours of video recordings from 50 healthy controls and 20 AD patients, focusing on conversational speech. However, its smaller sample size and the lack of detailed demographic backgrounds are notable limitations compared to DementiaBank. Our enhancement on DementiaBank focuses on improving consistency and representation to address the shortcomings found in datasets like AZTIAHO and aims to boost the reliability of research outcomes by providing a more comprehensive and demographically varied dataset.

Sample Characteristics and Representation

The dataset is curated by organizing and extracting information across different types of files (such as audio recordings, transcripts text, Excel sheets, and documents). We first match the structured information stored in Excel sheets and documents with the audio recordings using the participants' ID as the key and remove the participants whose audios or structured data are missing. Participants with meaningless values (e.g. NA and negative values) are also removed as it indicates the participants did not complete the interviews or assessment tests. In the final analysis, variables that were unique to the corpus were also filtered such as the "Cinderella transcribed" variable corresponding to the Baycrest corpus, as our goal is to create a generalized dataset that represents a more generic population.

The post-curated dataset includes basic information such as the participant's age, gender, and cognitive status, and some specific variables for each corpus (as Table 1 depicted). It ensures a diverse population representation and mitigates bias through sheer numbers and a broad geographic range by combining corpora across varies locations. Our sample comprises a diverse range of age groups from 40 to 85 years, allowing for a comprehensive examination of AD/ADRD across various life stages. Data collection encompassed diverse regions, spanning states and countries such as Canada, China, and Spain.

Participants were recruited and randomly sampled with no compensation through online outreach by a university or as patients in the hospital affiliated with the corpus. The primary inclusion criterion across most corpora was age, given the prevalence of AD/ADRD in the elderly. And some corpora initiate assessments with MoCA or MMSE to determine the need for further testing. Participant race disclosure was infrequent. For instance, the Ye corpus, conducted in China, exclusively included participants of Asian descent, rendering race disclosure inconsequential.

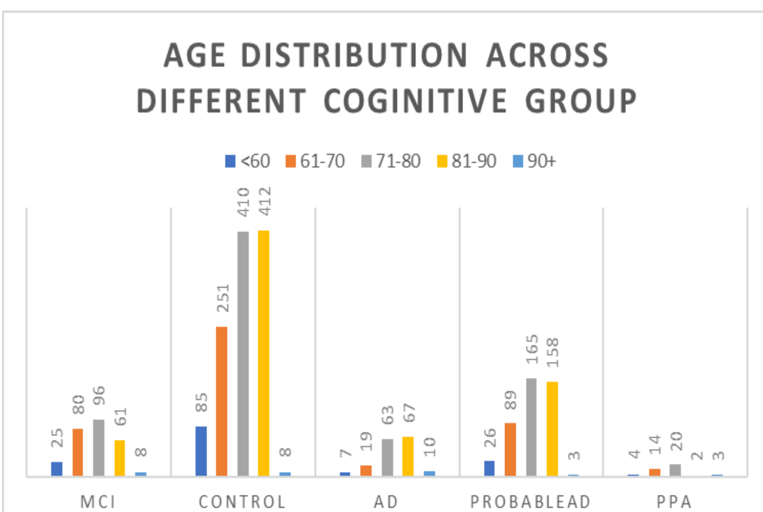

Figure 1 Age Distribution Across Different Cognitive Groups

Studies have shown that the risk of developing AD/ADRD increases as people age, and women seems to be disproportionately affected by this disease [22, 23]. After refining the dataset, Figure 1 and Figure 2 reveal trends in age and gender among the participants. The average age of the total participant pool is 75 with a high standard deviation as most participants were around the age of 81-83. The control group's average age hovers around 71, primarily falling below 70. The probable AD group's average age lies between 70-75, with ages ranging from 58 to 75. The AD group's average is closer to 78, with a distribution largely between 75 to 80 years.

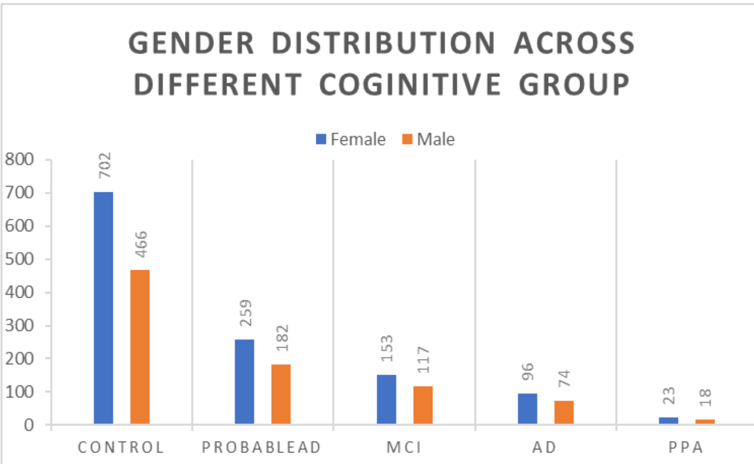

Figure 2 Gender Distribution Across Different Cognitive Group

In terms of gender, the dataset indicates a higher prevalence of females across all cognitive categories, highlighting the importance of exploring gender-related differences in cognitive decline. Notably, males with cognitive decline constitute 45% of their group, while females represent 43%. It's important to recognize that such figures only suggest correlations, not causation, due to the gender imbalance and lack of data on participants' lifestyles.

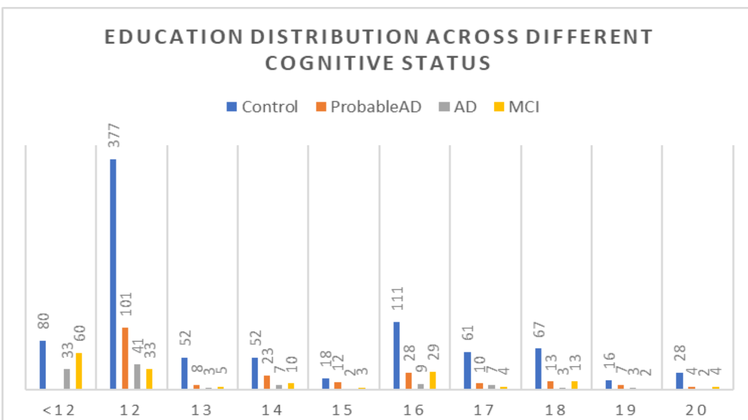

Figure 3 Education Distribution Across Different Cognitive Groups

Another factor would be the amount of education the participants received and their occupation. Individuals with lower income and education levels may face challenges in accessing healthcare, information, and support services [22, 23]. Regularly performing mental and cognitive tasks, whether through occupation or education, has been shown to decrease a participant's mental decline. As Figure 3 shows, participants tend to have lower education level and most of them are concentrated in education 11-12 and 16-18 levels. If we consider the percentage of participants, we can see that those with less education (11-12) are at greater risk of AD with the percentage averaging around 50%. For the participants with a higher education count, the pool of AD and Probable AD only makes up 30% (16-18). (Note: Not all corpus disclosed the education data)

Usability

Access to the data in the DementiaBank database is password protected and restricted to members of the DementiaBank consortium group. This ensures that the data is used ethically and responsibly, adhering to privacy and consent guidelines. For researchers who wish to access the DementiaBank data, the process typically involves applying and obtaining approval, which maintains a controlled environment for data use. This structure prioritizes the confidentiality and privacy of the data subjects, ensuring that the information is used solely for legitimate research purposes. We have applied and successfully got access to DementiaBank for this challenge, and plan to share the curated dataset with the management group, making it unified, protected and accessible for use.

The DementiaBank dataset has been curated, transforming it into a structured and unified resource that is ideal for advanced research into AD/ADRD. The curation process, including well-organized demographic, geographic, and clinical information, has ensured that the dataset is not just a collection of raw data but a coherent and systematically organized repository, which is ready-for-use and for various analytical approaches, like ML techniques and in-depth linguistic analyses. As a result, the dataset now facilitates efficient insight extraction and supports a more streamlined research workflow.

Considerations for further preparation before delivery include establishing a data dictionary to provide comprehensive guidance, detailing the structure, definitions, and constraints, and store the data in unified and common format to meet the requirements of various analytical tools. Additionally, the possibility of further anonymization to adhere to privacy standards may be necessary, ensuring that all data usage complies with ethical guidelines. Following the curation of the DementiaBank dataset, we intend to incorporate both audio files and de-identified demographic information for each participant. This curated data will encompass general attributes such as age, gender, and health status. However, to safeguard individual privacy, each participant will be represented using a de-identified ID. We will conscientiously exclude specific locational details, addresses, and detailed cognitive assessment information to prevent any possibility of personal identification.

In our commitment to protect individual privacy, we will adopt rigorous data anonymization techniques. This involves the careful removal and alteration of any potential personal identifiers in the dataset. Furthermore, access to this refined dataset will be strictly limited to authorized researchers who have committed to adhering to established ethical guidelines and privacy laws. These steps are critical in maintaining the integrity of the research while ensuring the utmost respect for the privacy and confidentiality of all participants involved.

The raw data we used for this study is stored in a password-protected database, access to which is restricted to members of the DementiaBank consortium group. The following are the contributors of the data cited across various corpuses (data - owner format) (only show the corpus included in our curation):

- Baycrest - Jed Meltzer
- Delaware - Alyssa Lanzi
- Lu - Maximillian Lu
- Pitt - Francois Boller and James Becker
- VAS - Xiaohui Liang
- WLS - Carol Roan
- Hopkins - Argye Hillis
- Ye - Zheng Ye
- Ivanova - Olga Ivanova

The aforementioned data is available for verification on DementiaBank. No copyright, irb-related, or consent-related constraints were met.

Team Information

Team Captain: Wenyao Xu (permanent resident of the United States)

Dr. Xu is a Professor and Associate Department Chair of Computer Science and Engineering Department in the State University of New York (SUNY) at Buffalo, where he founds and directs the ESC (Embedded Sensing and Computing) Group. He has published over 200 technical papers, co-authored 2 books and is a named inventor on many international and U.S. patents. He received the Ph.D. degree from University of California, Los Angeles (UCLA). He received both M.S. degree and B.S. degree from Zhejiang University (both with honor), China.

He focuses on exploring novel Embedded Sensing and Computing technologies to build up innovative Internet-of-Things (IoT) systems for high-impact real-world applications in the fields of healthcare, security and advanced manufacturing. Results has been published in peer-reviewed top research venues across multiple disciplines, including Computer Science (e.g., ACM SenSys, MobiSys, MobiCom, UbiComp, ASPLOS, ISCA, HPCA, Oakland, NDSS and CCS), Biomedical Engineering (e.g., IEEE TBME, TBioCAS, and JBHI), and Medicine (e.g., LANCET, Advanced Science).

Team Leader: Wei Bo (Female)

Wei is a third-year Ph.D. student of the Computer Science and Engineering Department in the State University of New York (SUNY) at Buffalo, under the supervision of Dr. Xu. She received M.S. degree from Washington University in Saint Louis, US and B.S. degree from Shanghai University of Finance and Economics, China.

She is motivated and combines hands-on analytical experience with research interests on smart health, embedded/mobile sensing, and related intelligent analytics/computing. She has published three papers (on GSA, IEEE TMI, and Elsevier SH) and is professional in how data and analysis inform practical applications to drive overall objective achievements in the engineering field.

Team Member: Harry Lin

Harry is a junior student of the Computer Science and Engineering Department in the State University of New York (SUNY) at Buffalo alongside an interest in pursuing a M.S. degree.

He is at the commencement of his career path and fueled by a resolute determination and a fervent eagerness to acquire new knowledge and embrace novel experiences, he has cultivated a keen interest in research. Beyond his academic pursuits, he dedicates his off-time to self-improvement, demonstrating a commitment to enhancing not only his scholarly abilities but also his overall personal development.

Team Member: Johnson Schwede

Johnson is a first-year undergraduate student majoring in computer science at Middlebury College. He received his high school degree from Nichols School located in Buffalo, New York.

He is a diligent and innovative student who is interested in conducting research in health-oriented subjects. He has previously conducted research in collaboration with the Department of Surgery at the State University of New York (SUNY) at Buffalo, Roswell Park Comprehensive Cancer Center, and Nichols School about the effects of antibiotics on long-term survival and recurrence free survival in non-small cell lung cancer lobectomy patients.

Reference

[1] “Dementia.” World Health Organization, March 15, 2023. https://www.who.int/news-room/fact-sheets/detail/dementia.
[2] “Alzheimer’s Disease Fact Sheet.” National Institute on Aging, April 5, 2023. https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet.
[3] DrivenData. “Prepare Challenge: Data for Early Prediction (Phase 1).” PREPARE: Pioneering Research for Early Prediction of Alzheimer’s and Related Dementias EUREKA Challenge, n.d. https://www.drivendata.org/competitions/253/competition-nih-alzheimers-adrd-1/.
[4] Vigo, Inês, Luis Coelho, and Sara Reis. "Speech-and language-based classification of Alzheimer’s disease: a systematic review." Bioengineering 9, no. 1 (2022): 27.
[5] DementiaBank, August 30, 2017. https://dementia.talkbank.org/.
[6] Haider, Fasih, Sofia De La Fuente, and Saturnino Luz. "An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech." IEEE Journal of Selected Topics in Signal Processing 14, no. 2 (2019): 272-281.
[7] Meilán, Juan José G., Francisco Martínez-Sánchez, Juan Carro, Dolores E. López, Lymarie Millian-Morell, and José M. Arana. "Speech in Alzheimer's disease: can temporal and acoustic parameters discriminate dementia?." Dementia and geriatric cognitive disorders 37, no. 5-6 (2014): 327-334.
[8] Kumar, M. Rupesh, Susmitha Vekkot, S. Lalitha, Deepa Gupta, Varasiddhi Jayasuryaa Govindraj, Kamran Shaukat, Yousef Ajami Alotaibi, and Mohammed Zakariah. "Dementia detection from speech using machine learning and deep learning architectures." Sensors 22, no. 23 (2022): 9311.
[9] Herd, Pamela, Deborah Carr, and Carol Roan. "Cohort profile: Wisconsin longitudinal study (WLS)." International journal of epidemiology 43, no. 1 (2014): 34-41.
[10] Shao, Zeshu, Esther Janse, Karina Visser, and Antje S. Meyer. "What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults." Frontiers in psychology 5 (2014): 772.
[11] Becker, James T., François Boiler, Oscar L. Lopez, Judith Saxton, and Karen L. McGonigle. "The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis." Archives of neurology 51, no. 6 (1994): 585-594.
[12] Mueller, Kimberly D., Bruce Hermann, Jonilda Mecollari, and Lyn S. Turkstra. "Connected speech and language in mild cognitive impairment and Alzheimer’s disease: A review of picture description tasks." Journal of clinical and experimental neuropsychology 40, no. 9 (2018): 917-939.
[13] Baek, Min J., Hyun J. Kim, Hui J. Ryu, Seoung H. Lee, Seol H. Han, Hae R. Na, YoungHee Chang, Jean Y. Chey, and SangYun Kim. "The usefulness of the story recall test in patients with mild cognitive impairment and Alzheimer's disease." Aging, Neuropsychology, and Cognition 18, no. 2 (2011): 214-229.
[14] Knopman, David S., and Soren Ryberg. "A verbal memory test with high predictive accuracy for dementia of the Alzheimer type." Archives of neurology 46, no. 2 (1989): 141-145.
[15] Julayanont, Parunyou, and Ziad S. Nasreddine. "Montreal Cognitive Assessment (MoCA): concept and clinical review." Cognitive screening instruments: A practical approach (2017): 139-195.
[16] Arevalo-Rodriguez, Ingrid, Nadja Smailagic, Marta Roqué i Figuls, Agustín Ciapponi, Erick Sanchez-Perez, Antri Giannakou, Olga L. Pedraza, Xavier Bonfill Cosp, and Sarah Cullum. "Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI)." Cochrane Database of Systematic Reviews 3 (2015).
[17] Werheid, Katja, Christian Hoppe, Angelika Thöne, Ulrich Müller, Martina Müngersdorf, and D. Yves von Cramon. "The Adaptive Digit Ordering Test Clinical application, reliability, and validity

of a verbal working memory test." Archives of Clinical Neuropsychology 17, no. 6 (2002): 547-565.
[18] “Creative Commons.” CC BY-NC-SA 3.0 Deed | Attribution-NonCommercial-ShareAlike 3.0 Unported | Creative Commons, n.d. https://creativecommons.org/licenses/by-nc-sa/3.0/.
[19] “Basic Rules for Data Usage.” TalkBank, n.d. https://talkbank.org/share/rules.html.
[20] Eyben, Florian, Klaus R. Scherer, Björn W. Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y. Devillers et al. "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing." IEEE transactions on affective computing 7, no. 2 (2015): 190-202.
[21] López-de-Ipiña, Karmele, Jesus-Bernardino Alonso, Carlos Manuel Travieso, Jordi Solé-Casals, Harkaitz Egiraun, Marcos Faundez-Zanuy, Aitzol Ezeiza et al. "On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis." Sensors 13, no. 5 (2013): 6730-6745.
[22] Gilmore-Bykovskyi, Andrea L., Yuanyuan Jin, Carey Gleason, Susan Flowers-Benton, Laura M. Block, Peggye Dilworth-Anderson, Lisa L. Barnes, Manish N. Shah, and Megan Zuelsdorff. "Recruitment and retention of underrepresented populations in Alzheimer's disease research: a systematic review." Alzheimer's & Dementia: Translational Research & Clinical Interventions 5 (2019): 751-770.
[23] “Addressing Diversity in Alzheimer’s Clinical Trials: AAIC 2021.” AAIC, July 29, 2021. https://aaic.alz.org/releases_2021/clinical-trial-diversity.asp.