1. Please provide your preferred information for use in announcing the winners of the competition:
    a. Name (first and last name or first name and last initial):
        i. Julie Fennell, Ph.D.
        ii. David Kaniss
    b. Organization: Engineering Dynamics (HQ in Silver Spring, Maryland)
    c. Hometown:
        i. Julie lives in Silver Spring, Maryland
        ii. Dave lives in Silver Spring, Maryland
    d. A recent picture of yourself or digital avatar (feel free to attach separately):
2. Social handle or URL (optional):
3. Who are you (mini-bio) and what do you do professionally?
    a. Julie is a sociologist and author of *Violated: Sexual Consent and Assault in the 21st Century* (2024). Her current research focuses on psychological disadvantages experienced by sexual minority groups, which is how she became drawn to the PREPARE project.
    b. Dave Kaniss is a senior Systems Engineer, specializing in long-term sustainability of complex, integrated systems. His experience covers the entire developmental lifecycle of such systems, ensuring they are safe, cost-effective, reliable, and user-friendly.
4. What motivated you to compete in this challenge?
    a. Julie's current research focuses on psychological disadvantages experienced by sexual minority groups, which is how she became drawn to the PREPARE project.
    b. Dave is deeply committed to using AI to help people, addressing challenges that range from everyday practicalities to more light-hearted endeavors. Using AI to identify early stages of Alzheimer's is crucial, as it opens the door to timely interventions, potentially altering the course of the disease and significantly improving the quality of life for those affected.
5. High level summary of your dataset: the data source, target, predictors, sample size and use for early, inclusive prediction of AD/ADRD.
    a. Our dataset uses recodes of the National Health Interview Survey (NHIS), compiled and harmonized by IPUMS (Integrated Public Use Microdata Series), which is a largescale nationally representative survey of the non-institutionalized United States population, ages 18-85. Relevant years currently available are from 2010-2022, although not all major variables are available for all years. The average annual sample is around 100,000 people in 45,000 households. Key dependent variables are DEMENTIAEV ("have you ever been told by a health professional that you have dementia, including alzheimer's") from 2019-22; LAMEMORCON ("have you ever

had difficulty remembering, concentrating, or both?"), which can be focused solely on respondents who had difficulty remembering, from 2010-22; LAMEMCONDIF, which asks for level of difficulty remembering for the same years; LAMEMDIFOFT, which asks for frequency of difficulty remembering for the same years.

6. What are two or three unique strengths of this dataset or type of data for early, inclusive prediction of AD/ADRD?

    a. The vast sample size of the NHIS means that even extremely small groups can potentially be targeted when data are pooled, including immigrants based on years in the U.S., sexual minority groups, as well as people with specific health conditions like diabetes and asthma. Moreover, the scaled responses in terms of difficulty remembering allow for more nuance in predicting memory loss than black-and-white measures of cognitive decline do.

7. Did you use any tools or resources for developing your submission (e.g., to find a dataset, or explore the contents of a public dataset)?
    a. We are grateful to the dedicated employees at IPUMS for all their hard work in maintaining and distributing this massive data set, and particularly to Julia Drew, who was extremely personally helpful in navigating this data.

8. Were there any data types or sources that you explored but didn't fit for this challenge?

    a. N/A

9. How would you improve or enrich this dataset if you had access to a big research team and an unlimited budget?

    a. With an unlimited budget, like many big data set researchers, we share the dream of combining information from a few nationally representative surveys on the few years they have overlapped samples, and the NHIS is one of the major links in those big national survey sets. Moreover, there is confidential data in the NHIS that is not widely available to researchers that we would love to be able to have indefinite access to, but currently it is extremely restricted due to understandable confidentiality concerns.