**Analyzing the Impact of Financial Stability and Family Structure on Elder Health and Longevity: A Machine Learning Approach Using the RAND HRS Longitudinal File 2020**

Yuzhen Zhou, Zeyuan Pang

**Part 1. Executive Summary**

The primary goal of our research is to analyze the impact of financial stability and family structure on the health outcomes. This investigation is crucial as nations globally grapple with the ramifications of an aging demographic, including heightened demands on healthcare systems, pension sustainability, and the overall welfare of the elderly population. Our study aims to provide evidence-based insights to guide policymaking in healthcare, social security, and family support initiatives, thereby facilitating more effective strategies to aid older adults.

For this analysis, we utilize the RAND HRS Longitudinal File 2020, which encompasses 15 waves of interview data collected over two decades. This comprehensive dataset is invaluable for research on health, family dynamics, retirement planning, employment history, and includes imputations for income, assets, and healthcare spending.

By examining the connections between economic status, family structure, and the health of the elderly, we aim to recommend targeted interventions that could improve life quality, reduce medical costs, and promote the sustainability of aging populations.

This project is of particular importance as it confronts a pressing challenge faced by East Asia: its rapidly aging population. This demographic transformation poses significant challenges for the social and economic progress of these countries, necessitating innovative approaches to ensure the well-being of the elderly and their families in a changing societal landscape.

**Part 2. Data Exploration and Preprocessing**

RAND HRS Longitudinal File 2020 is a huge dataset. It took weeks to read and understand the document and organize the data from different waves into a single dataset.

After reading the document, several variables of interest were selected for the initial analysis. The response variables are “Self-rated Health Level”, “Cognitive Level”, and “Mental Status Level”