



**Application number/Title:** 19039 - Big Data Integration Methods and Applications to UK Biobank Data

**Applicant PI:** Professor Hulin Wu

**Applicant institution:** University of Texas (UT Health), Biostatistics, 1200 Pressler Street, RAS E809, Houston, TX 77030, United States

**Keywords provided by the Applicant PI to describe the research project:**

Big data

**Application Lay Summary:**

**1a:** Aim 1: Develop novel high-dimensional modeling approaches to integrate genotyping, phenotype and biomarker data for the prediction of death and cancer outcomes.

Aim 2: Develop novel network approaches to integrate genotyping and multimodal imaging data (including MRI and DXA scans) to predict death and cancer outcomes.

Aim 3: Develop novel time course modeling approaches to model accelerometry, dietary and behavioral data to predict death and cancer outcomes. Raw accelerometer data will allow us to understand the true signal and the noises, which is important for the statistical methodology.

**1b:** We expect that our new approaches will result in important health science findings from the UK Biobank data, since we expect to integrate genotype, biomarker, dietary, behavioral and imaging data (including MRI and DXA) to predict death and cancer outcomes. We will follow the UK Biobank policy to share our methodologies and research results with the general public.

We intend to use imaging data (including MRI and DXA) to define possible novel variables and also to verify or derive new approaches for data pre-processing. Imaging data will be used since these are potential predictors of chronic diseases and cancer.

**1c:** We will employ big data approaches for the preparation and analysis of the data provided by the UK Biobank. These approaches will be based on standard statistical methods and techniques. Which methods are used in each

approach can only be determined from the characteristics of available data. This project's output includes novel statistical and computational approaches to integrate the variety of complex data in the UK Biobank for outcome predictions. From the use of these novel approaches we expect to obtain novel predictions of diagnostics of death and cancer.

**1d:** Full Cohort