We aim to discover novel clues in large-scale electronic health records (EHR) to prevent the onset of Alzheimer's disease.

Identification of Factors Associated with Alzheimer's Disease Diagnosis Over Complex Large-scale Longitudinal Health Data
Beth Shaaban, Helmet Karim, Scott Malec, Sanya Bathla Taneja, Maria Ly, Jonathan C. Silverstein, Steven Albert, Paul Munro, Richard D. Boyce

Motivation
Risk of Alzheimer's disease (AD) doubles every 5 years after age 65. An estimated 14 million Americans will have AD by 2050. With no disease modifying treatment or prevention in sight, we aim to:
- Identify factors associated with pre-clinical AD i.e. 10 or more years prior to disease diagnosis using electronic health records.
- Discover novel causal associations for AD using graphical causal methods.

Project Description
1. Cases - patients with ICD-10 AD diagnosis since 2016 and with UPMC visits 10+ years prior to their diagnosis.
2. Controls (8 controls to 1 case) - patients matched on age, sex, and other factors who were not diagnosed with AD but do have UPMC visits 10+ years prior.
3. Case-control & machine learning analyses to identify early markers of AD.
4. Knowledge graphs and graphical causal methods to address confounding and suggest mechanisms.

Context
- Type of analysis
  First analysis of its kind:
  - Associations 10+ prior to a diagnosis of Alzheimer's disease
  - Inpatient + outpatient data
  - Complete EHR data - medical history, medications, lab results, family history, procedures, demographics, and vitals.
  - Integrates detailed control of confounders identified through both typical literature review and machine reading of the literature
- Scale of analysis
  > 9,000 cases and >72,000 controls
- Team
  Multidisciplinary members from three schools (Medicine, Public Health, Computing & Information)

Project Deliverables
- Data mart of UPMC EHR data with both cases and controls
- Case-control and machine learning analyses for AD and associated factors
- Preliminary results on causal associations through knowledge graphs
- Abstracts and manuscripts with results
- National Institute of Ageing (NIA) R01 grant proposal for long-term funding

Potential Impact
Provide novel hypotheses to advance research on prevention of AD
- Identify potential disease-course-altering markers for further investigation
- Long-term research plans (using follow-on funding) will generate much-needed disease trajectories & mid-level mechanistic models of great value to AD research community
- A new data mart for use by the wider community of AD researchers

Collaborators
Dr. Arthur Levine
Michelle Kienholz
Dr. Howard Aizenstein
Dr. William Klunk
Brian McLay and the R3 team