NACC Project #2012-JI-04 (secondary analysis of NACC data)

Missing Data and the Identification of Optimal Predictors of Conversion to AD

Junior Investigator

Maritza Dowling, MS, PhD, University of Wisconsin

ADC Director

Sanjay Asthana, MD, FRCP

Project Work Dates

2012–2013

Project Description

A major interest in ongoing Alzheimer’s disease (AD) research is on identifying biomarkers that best predict which individuals will progress to AD. However, the accuracy in the identification of such discriminative predictors of disease progression may be substantially influenced by missing data; a key challenge primarily in aging research. Missing data have the potential to produce biased results and reduce the accuracy of discriminating between those who convert to AD from those who do not. The primary goal of this project is to perform a comprehensive analysis of the impact of different approaches for handling missing data on the mis-classification probabilities of a distal AD diagnosis and the identification of relevant predictors of conversion to AD.

Hypothesis: An appropriate selection of methodology for handling missing data results in significant improvements in the sensitivity and specificity of the prediction model for discriminating converters from non-converters in the MCI group. Further, the choice of missing data treatment will affect the relative importance of predictors of conversion to AD.

Specific Aims: (1) Conduct a systematic assessment of the a) amount, level, and pattern of missingness in potential predictors and outcome of interest and b) assumptions about the observed missing data mechanism that are sensible to make given the characteristics of the data. (2) Use information information from Aim 1 to implement five approaches for handling missingness. (3) Apply novel variable selection methodology to reduce dimensionality, and (4) Compare the predictive accuracy across methods.

Methods: We will estimate and compare the performance of three expert-recommended multiple imputation techniques and two likelihood-based approaches for handling missing data. Dimensionality reduction of the predictor set and estimation of parameters will utilize a novel penalized likelihood-based method.

Significance: This project is innovative in that it presents a comprehensive comparison of state of the art techniques to treat missing data and their effect on model accuracy in predicting conversion to AD.

Contact Information

For more information about this study, please contact:

Maritza Dowling, MS, PhD
University of Wisconsin
email: nmdowlin@biostat.wisc.edu
ADC website: www.wcmp.wisc.edu