Analytic Methods for Determining Multimodal Biomarkers for Parkinson’s Disease

F. DuBois Bowman

Department of Biostatistics and Bioinformatics
Emory University, Atlanta, GA, 30322

NINDS PDBP Kick-off Meeting
Arlington, VA

Nov 13, 2012
Acknowledgements

Research Team

- Emory University and CBIS Collaborators:
  - Lijun Zhang, PhD, *Biostatistics and Bioinformatics*
  - Jian Kang, PhD, *Biostatistics and Bioinformatics*
  - Ying Guo, PhD, *Biostatistics and Bioinformatics*
  - Xiaoping Hu, PhD, *Biomedical Engineering*
  - Anthony Pileggi, *Biostatistics and Bioinformatics*

- Kaiser Permanente Georgia.
  - Daniel Huddleston, MD, *Neurology at Emory*
  - Michele Marcus, PhD, *Epidemiology at Emory*
General Overview

- Our project seeks to develop statistical tools to identify multimodal biomarkers of Parkinson’s disease (PD).
  - Establish a PD risk profile to identify subjects to populate future clinical trials (e.g. assessing neuroprotective treatments).
  - Examine multimodal imaging (MRI-based), biologic, and clinical candidate biomarkers.
    - Hypothesis driven biomarkers
    - A massive number of exploratory biomarkers

- Develop software for the PD biomarker research community to help implement our planned statistical tools.

- Make our data publicly available to the research community.
Motivation

The optimal window for neuroprotection in PD is during the **premotor period** before most neuronal death occurs.

Neuropathological evidence suggests that **locus coeruleus (LC)** may be involved with PD pathology before SNpc (Braak Hypothesis).


**Neurobiology of Aging** 24 (2003):197-211.

[dm: dorsal motor nucleus of the glossopharyngeal and vagal nerves; co: locus coeruleus (pons); sn: substantia nigra (midbrain); mc: anterior temporal mesocortex; hc: high order sensory association areas and prefrontal fields; fc: first order sensory association areas, premotor areas, and primary sensory and motor fields.]
The catecholamine nuclei, substantia nigra (pars compacta) and locus coeruleus, degenerate in Parkinsons disease.

*NeuroReport* 9:1649-1654.
Motivation

Locus Coeruleus Cell Loss


[AD (n=86); PD (n=19); HC (n=13)]

Motivation

Why MRI?

- **Safe, non-invasive**, longitudinal assessment in living subjects.
- **Multimodality**: can acquire multiple types of information in a single scanning session.
  - MRI, fMRI, DTI, NM-MRI, CSI
- **Translational application**: MRI is widely available and familiar technology in clinical settings.
- **Inexpensive** (relative to PET/SPECT) and no radionuclide exposure.
Motivation

Neuromelanin MRI reveals *decreased contrast* with neurodegeneration.
Motivation

DTI in Substantia Nigra

High-resolution diffusion tensor imaging in the substantia nigra of de novo Parkinson disease

Figure 1 - Procedure used to draw regions of interest in the substantia nigra

Figure 3 - Differences between individual subjects

(A) Statistical analysis of the fractional anisotropy values from rostral (red), middle (green), and caudal (blue) regions of interest. The black symbol represents the mean, and the error bars represent the standard error of the mean.

F. D. Bowman (Emory University) Determining Multimodal PD Biomarkers PDBP Kick-off Meeting 10 / 19
Objectives

Multimodal biomarker detection

- Numerous findings suggest links between PD and single genetic, imaging, and biologic factors.
  - Many of these are non-specific or insensitive.
  - Single modality biomarkers may not fully address the complexity of PD.

- We regard PD as a complex, systems-level, multi-dimensional disorder with discrete, but functionally integrated manifestations.

- We will develop methods to define multimodal PD biomarkers from a massive number of hypothesis driven and exploratory candidate markers.
Specific Aims

**Aim 1:** To Develop new statistical techniques to reveal multimodal biomarkers for PD including imaging, clinical, and biologic variables.

- Collect MRI, fMRI, DTI, NM-MRI, CSI imaging data along with genetic, clinical, and CSF-based measures.
- NM-MRI uses an *in-house* optimized pulse sequence, which captures both the LC and SN in $\approx 16$ minute scan, and we apply a contour segmentation algorithm to generate LC volume estimates.
- Develop statistical model for high-dimensional data to pool strength across multiple data modalities for classifying subgroup membership (e.g. PD or HC).
  - $\ell_p(\beta, \lambda) = -\ell(\beta) + \sum_k \lambda_k P_k(\beta_k)$
  - $\ell(\beta) = \sum_i [y_i \log \pi_i + (1 - y_i) \log (1 - \pi_i)]$
- Variable selection using modality-specific penalties
- Model development, training, testing, and validation
- Apply our developed tools to data from imaging-based studies.
Aim 2: To identify prediagnosis clinical predictors of PD from a massive database obtained from an integrated, closed cohort healthcare system.

- Utilize Kaiser Permanente Georgia patient database to select PD patients (with medical history at KP) and healthy controls.
- Data will include patient subgroup (PD or not) along with diagnoses, medication history, and lab results.
- Develop regularized logistic predictive model to select significant clinical factors for predicting the probability of PD.
  - Cross-sectional clinical factors
  - Longitudinal clinical factors
- Generate risk score for development of PD.
  - Eligibility criterion for enrollment into future studies (e.g. clinical trials of neuroprotective treatments).
Specific Aims

**Aim 3:** To develop software equipped with a friendly graphical user interface to implement the multimodal biomarker detection methods.

- **Data management** system for storing, sharing, and securing data.
- **Data integration** component for fusing the different multimodal data and implementing preprocessing algorithms
- **Analysis** GUI for biomarker exploration, which supports execution of the approaches proposed in Aims 1 and 2.
Data

Multimodal Imaging-based Studies

- The data will include 81 subjects across three studies
  - 38 Parkinson’s disease patients
  - 32 Healthy control subjects
  - 11 Alzheimers disease patients

- Potential Biomarkers
  - Imaging
  - Genetic
  - Neurocognitive Testing
  - Questionnaire-Derived Scores
  - Clinical
  - CSF Neuroinflammation
  - CSF Catecholamine Metabolites
Clinical Database from Kaiser Permanente Georgia Healthcare System

- The KP Georgia system includes over 235,000 patients
- Approximately **530 Parkinson’s disease patients** with at least one year of clinical observation prior to PD diagnosis
- Clinical chart database (Epic) allows data pulls for research.
- Potential PD Predictors
  - **ICD9 Codes**: depression, anxiety, constipation, anosmia, orthostatic hypotension, REM sleep behavior disorder
  - **Medications** used to treat these disorders: antidepressants, anxiolytics, constipation medications, fludrocortisone/midodrine, clonazepam
  - **Labs**: Anemia, EKG, electrolytes, liver function, etc.
- Very inexpensive to identify large numbers of potential pre-motor PD subjects.
Selected features from multiple imaging modalities.

- Blue represents PD patients and red represents HC subjects.
- Networks: FC; Local activity: ALFF; Volumetric: VBM; Chemical shift imaging: RLC and LLC.

[NM-MRI Estimated Locus Coeruleus Volume. Controls: N=6; PD: N=9].
Multimodal classification of PD patients versus HCs

- Networks: FC
- Local activity: ALFF
- Volumetric: VBM
- Chemical shift imaging: RLC and LLC
- Multimodal dataset (blue)

Number of predictors and the AUC are given.
Thank you!

Emory University and CBIS Collaborators:
- Lijun Zhang, PhD, *Biostatistics and Bioinformatics*
- Jian Kang, PhD, *Biostatistics and Bioinformatics*
- Ying Guo, PhD, *Biostatistics and Bioinformatics*
- Xiaoping Hu, PhD, *Biomedical Engineering*
- Anthony Pileggi, PhD, *Biostatistics and Bioinformatics*

Kaiser Permanente Georgia.
- Daniel Huddleston, MD, *Also Neurology at Emory*
- Michele Marcus, PhD, *Also Epidemiology at Emory*