**What is PPMI?**

The Parkinson’s Progression Markers Initiative (PPMI) is a landmark observational clinical study to comprehensively evaluate cohorts of significant interest using advanced imaging, biologic sampling and clinical and behavioral assessments to identify biomarkers of Parkinson’s disease progression. PPMI is taking place at clinical sites in the United States, Europe, Israel, and Australia. Data and samples acquired from study participants will enable the development of a comprehensive Parkinson’s database and biorepository, which is currently available to the scientific community to conduct field-changing research.

What does the dataset contain?

Dataset in PPMI contains the following:

* **Clinical Markers**
  + Cognition - Montreal Cognitive Assessment (MoCA)
  + Behavioral
    - Depression - Geriatric Depression Scale (GDS-15)
    - Anxiety - State Trait Anxiety Inventory for Adults
    - ICS
  + Autonomic
    - Constipation
    - Bladder
    - Sexual
    - Cardiac
  + Olfaction - UPSIT
  + Sleep RBD (Epworth Sleepiness Scale & REM Sleep Behavior Questionnaire)
  + Motor Analysis
  + Speech (Semantic fluency)
  + Medications & Examinations (General & Neurological)
  + Medical and family history (including demographics)
  + Socio Economic Status
  + Household & Leisure Activity
  + Lumbar Puncture Sample
  + MDS-UPDRS scores (including Part III and Hoehn & Yahr)
  + Smoking and Alcohol Information
  + Hopkins Verbal Learning Test & Vital Signs
  + QUIP
* **Imaging-Phenotomics**
  + SPECT/PET - Depamine DAT, VMAT2
  + Amyloid
  + MRI-DTI, volumetrics [Diffusion Tensor Imaging]
  + Structural MRI Imaging
* **OMICS**
  + RNA Profiling
* **Genetics**
  + Synuclein, GBA etc
  + SNP Array Data
  + Exome Sequencing
* **Biologics**
  + Blood, CSF, Urine
  + Skin Biopsy
  + Lab Procedures

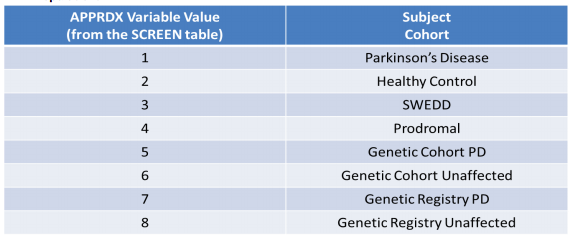
**Data Structure**

Currently there are 92 data tables (.csv files) comprising both clinical and lab data

Useful files to figure out the variables:

* Data Dictionary: lists all variables in each data table with a brief description o
* CRFs (Case Report Forms)
* Code List: lists every categorical variable with each possible response (CODE) and what each category means (DECODE) CODE = 1; DECODE = “Yes”
* PPMI Derived Variable Definitions and Score Calculations
* **PPMI Populations:**

The APPRDX in the screen table determines PPMI populations

****

* Finding enrolled subjects
  + Need two tables: SCREEN (Screening/Demographics) and RANDOM (Randomization table)
  + Enrolled subjects must appear in both tables
  + Enrolled subjects must also have a non-missing enrollment date (ENROLLDT from the RANDOM table)

**Critical Dataset Information:**

* Age of onset: Age of onset can be determined by looking at the Year of Diagnosis under the column heading MHDIAGYR in the General Medical History spreadsheet.
* CLINICAL\_EVENT and EVENT\_ID both represent the visit number. There is no difference between the two.
* Inclusion Exclusion Criteria: The Inclusion Exclusion csv provides the inclusion/exclusion information. the INEX1 column on the Inclusion/Exclusion worksheet, 0 means no and 1 means yes
* PD subjects classification: Classification of TD vs. PIGD can be derived based on the calculations included in the Derived Variable Definitions and Score Calculations document
  + TD = tremor dominant
  + PIGD = postural instability and gait disturbance

PMMI Data Stats:

* 3 Baseline files
  + Baseline Cognitive Data contains:
    - 618 Patients with 22 Attributes
    - Min age: 30 & Max age: 85 years, Mean: 62 years
    - Key Attributes: MoCA Scores, QUIP Info, Duration of Disease, Benton.Judgment.Of.Line.Orientation, Semantic.Fluency.Total.Score Letter.Number.Sequencing.Total.Raw.Score, Symbol.Digit.Modalities.Score
  + Baseline Data Summary contains:
    - 564 Patients with 22 Attributes
    - Key Attributes: MDS UPDRS, UPSIT Total Score, SBR, HNY
  + Baseline Specimen Collection contains:
    - 619 rows with 13 Attributes
    - Key Attributes: DNA, Serum, Plasma, Urine in ml
* 92 CSV Files contains detailed information provided from the cognitive, data summary and specimen files
  + Key Information Files
    - Data Dictionary (Data\_Dictionary.csv)
    - Biospecimen (Biospecimen\_Analysis\_Results.csv, Blood\_Chemistry\_\_\_Hematology.csv)
    - Cognitive Assessments (Cognitive\_Assessments.csv, Cognitive\_Categorization.csv, Montreal\_Cognitive\_Assessment\_\_MoCA\_.csv)
    - Clinical Diagnosis (Clinical\_Diagnosis\_and\_Management.csv, Diagnosistic\_Features.csv, Current\_Medical\_Conditions\_Log.csv, General\_Medical\_History.csv, General\_Physical\_Exam.csv, General\_Neurological\_Exam.csv, Primary\_Diagnosis.csv)
    - DAT Scans - (DATScan\_Analysis.csv, DaTscan\_Imaging.csv, DaTSCAN\_SPECT\_Visual\_Interpretation\_Assessment.csv)
    - Lab Samples & Procedures (Laboratory\_Procedures.csv, Laboratory\_Procedures\_with\_Elapsed\_Times.csv, Whole\_Blood\_Sample\_Collection.csv)
    - MDS UPDRS (Files starting with MDS)
    - Sleep (REM\_Sleep\_Disorder\_Questionnaire.csv,
    - SCREEN & Random (Randomization\_table.csv)
    - QUIP (Files starting with QUIP)
    - TAP (Files starting with TAP)
    - Vital Signs
    - PD Medication