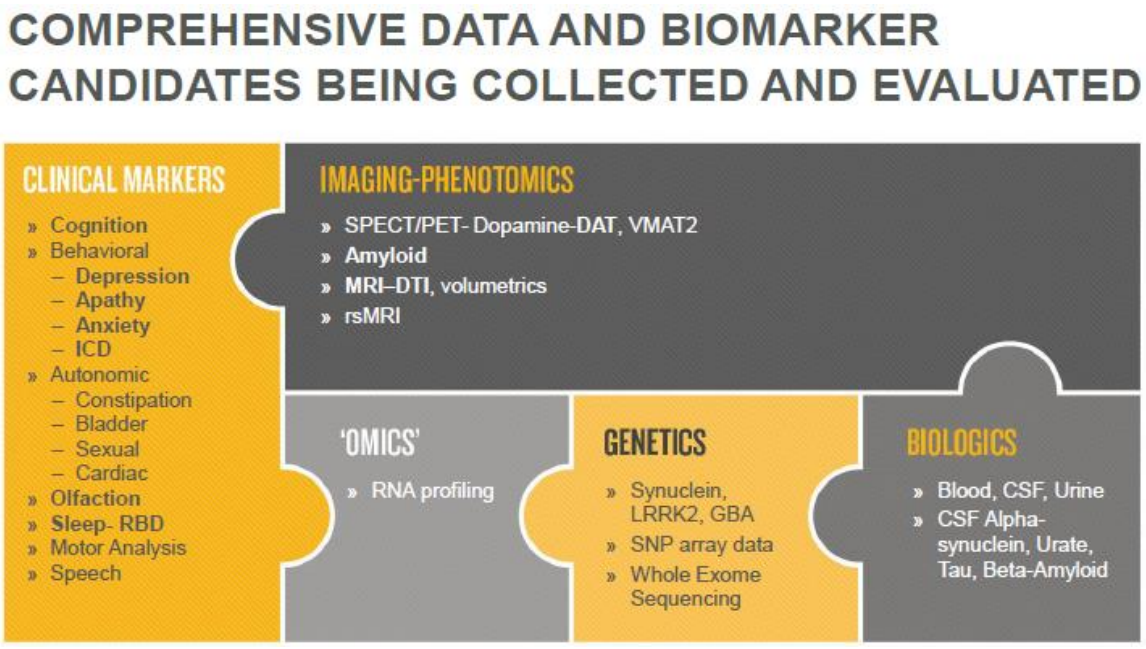
## PPMI Data Challenge – Dataset Overview and FAQ




**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.

| GROUP | Consented | Enrolled (n) | Withdrawn | Active | Complete |
|---|---|---|---|---|---|
| PD Subjects | 488 | **423** | 45 | **371** | **7** |
| Healthy Controls | 241 | **196** | 20 | **170** | **6** |
| SWEDD Subjects | 82 | **64** | 9 | **5** | **50** |
| Prodromal -Hyposmic | 118 | **26** | 2 | **24** | 0 |
| Prodromal-RBD | 96 | **39** | 0 | **39** | 0 |
| LRRK2 PD Cohort | 103 | **92** | 2 | **90** | 0 |
| LRRK2 UA Cohort | 94 | **85** | 2 | **83** | 0 |
| SNCA PD Cohort | 14 | **12** | 0 | **12** | 0 |
| SNCA UA Cohort | 4 | **4** | 0 | **4** | 0 |
| GBA PD Cohort | 18 | **17** | 0 | **17** | 0 |
| GBA UA Cohort | 26 | **21** | 0 | **21** | 0 |
| PD Registry | 120 | **114** | 3 | **111** | 0 |
| UA Registry | 127 | **120** | 0 | **120** | 0 |
| **TOTAL** | **1531** | **1213** | **83** | **1067** | **63** |

**PPMI Cohort Enrollment-2016**

This summary of cohort data to the left provides an overview of the various cohorts enrolled in the PPMI dataset. This dataset is a snapshot of the frozen dataset as of 4/4/2016.

**Data Structure**

- Currently there are 72 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
    - 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**

| APPRDX Variable Value (from the SCREEN table) | Subject Cohort |
|---|---|
| 1 | Parkinson's Disease |
| 2 | Healthy Control |
| 3 | SWEDD |
| 4 | Prodromal |
| 5 | Genetic Cohort PD |
| 6 | Genetic Cohort Unaffected |
| 7 | Genetic Registry PD |
| 8 | Genetic Registry Unaffected |

- 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)

**PPMI Dataset Frequently Asked Questions**

**Q:** In the Biospecimen data, some subjects have two measurements on the same test name per visit for CSF. Why? Which values should be used for analysis?
**A:** The CSF was run twice and we did get different results each time. The steering committee posted a note and a slide show explaining this and telling researchers how they should interpret/use the data: http://www.ppmi-info.org/access-data-specimens/.

**Q:** When a prodromal subject converts to PD, where is the data recorded? Do they now get enrolled in the PD cohort, restart at baseline, and follow the PD schedule?
**A:** For the PPMI prodromal subjects, a Diagnostic Questionnaire Form is completed at each visit. This data can be found in the data set called "Prodromal Diagnostic Questionnaire" under the Medical History. It has the PAG_NAME = "PRODDIAG" with the variable names of "PRIMDIAG".

**Q:** How do you interpret the Inclusion/Exclusion criteria spreadsheet?
**A:** You can view the Inclusion/Exclusion CRF, by clicking on Case Report Forms in the Research Documents & SOPs

section of the PPMI website. The numbers of the questions in the CRF will correspond to the INEX numbers in the Inclusion/Exclusion worksheet (i.e. INEX1 corresponds to question 1 on the Inclusion/Exclusion CRF, etc.). The Code List will allow you to determine what the numbers in the INEX columns mean (i.e. for the INEX1 column on the Inclusion/Exclusion worksheet, 0 means no and 1 means yes, etc.).

**Q:** How do I determine the age of onset for PD patients?
**A:** You can determine age of onset by looking at the Year of Diagnosis under the column heading MHDIAGYR in the General Medical History spreadsheet.

**Q:** How can I determine if PD subjects are tremor dominant vs. postural instability and gait disturbance?
**A:** Classification of TD vs. PIGD can be derived based on the calculations included in the Derived Variable Definitions and Score Calculations document.

**Q:** How can I access pre-processed imaging data?
**A:** Once you check off Pre-processed under IMAGE TYPES in the Search Options box on the left hand side of the Advanced Search (beta) page, you will be able to search for pre-processed images. Please note that the default setting is for this box to be unchecked.

**Q:** Why can't I view the imaging data using Chrome?
**A:** Unfortunately, Chrome no longer supports the plug-in needed to view the imaging data. Please try a different browser.

**Q:** I listed my co-investigators when I applied for access to the PPMI database, but now there are new co-investigators working on my project. How can I update this list?
**A:** Once you are logged into the LONI database, click My Account in the top right hand corner. In the Data Use section, click Update. From there, click the Co-Investigator tab and add the co-investigators. You can continue to update this section as needed.

**Q:** I understand that all PD patients are PD drug naïve at baseline, but how can I determine when subjects start taking PD medications?
**A:** Once you have logged into the LONI interface, click on the Download tab and then click Study Data to view the available data spreadsheets. On the menu of the left side of the screen, click Medical History and then click Use of PD medication to download the Use of PD Medication spreadsheet. A 1 in the PEDMEDYN column indicates that the subject had already started PD medications by that visit. The date the subject started PD medications can be found in the column titled PDMEDDT.

**Q:** Some data files use CLINICAL_EVENT and others use EVENT_ID. Is there a difference?
**A:** CLINICAL_EVENT and EVENT_ID both represent the visit number. There is no difference between the two.