Parkinson's Progression Markers Initiative

Guide to the PPMI Analytic Dataset

Background/Rationale: PPMI analytic dataset was created to enable data analysis to reflect *the most accurate current participant cohort and subgroup designation*. PPMI is an active on-going study that posts new information in real time. As such, the study will generate new clinical, genetic, imaging and biomarker data that may impact participant cohort/subgroup assignments that will impact data analysis. **Analytic dataset should be used for all PPMI data analyses. Analytic data set will be updated on regular basis and new version will be posted with the corresponding date of update.**

Key Definitions of cohorts and subgroups

Enrollment cohort and subgroups – defined by the participant allocation to the cohort and subgroup at enrollment based on the PPMI protocol eligibility definitions – see below. **The PPMI database at LONI is based on the enrollment cohorts.**

Cohort	Subgroup	LONI nomenclature	Analytic dataset nomenclature needed to define groups (refer to Excel, derived variables nomenclature does not yet exist in LONI)
Parkinson's disease	Sporadic	ENRLPD	CONPD*, CONLRRK2, CONGBA, CONSNCA
	Genetic / LRRK2	ENRLLRRK2	
	Genetic /GBA	ENRLGBA	
	SNCA/or rare mutations	ENRLSNCA	
Prodromal	RBD	ENRLRBD	CONRBD, CONHPSM, CONPROD, CONLRRK2, CONGBA, CONSNCA
	Hyposmia/ general risk	ENRLHPSM	
	Genetic / LRRK2	ENRLPROD, ENRLLRRK2	
	Genetic /GBA	ENRLPROD, ENRLGBA	
	SNCA/or rare mutations	ENRLPROD, ENRLSNCA	
Healthy controls	N/A	ENRLHC	CONHC
SWEDD (legacy)	N/A	ENRLSWEDD	CONSWEDD

***CON**=consensus

Analytic cohort and subgroups - The PPMI steering committee has recognized that ongoing collection of clinical, genetic, imaging and biomarker data may result in a change in the enrollment cohort and subgroup. Based on these data, participants may be removed from their enrollment cohorts or subgroups or may be re-categorized from their enrollment cohort and subgroups to different cohorts or subgroups. While these changes will affect only a small number of participants, this data driven change from enrollment cohort will optimize ongoing PPMI data analysis. **The revised enrollment cohort and subgroup assignment, now called the PPMI analytic dataset,** (including the current correct cohort and subgroup assignment for each participant) is in the attached Excel spreadsheet and should be used for all PPMI data analyses. The appropriate variables needed to define each analytic subgroup are provided in the table above. More specific details are provided in the attached Excel spreadsheet.

How to use analytic dataset

LONI dataset reflects enrollment cohorts and subgroups. Investigators who want to incorporate participants classification from the analytic dataset in their analysis should cross reference the enrollment and analytic datasets as presented in the Excel spreadsheet and use analytic dataset participant classification.

Examples

Analysis of the PD cohort

- should exclude participants labeled not-PD

- may exclude participants with genetic variants (depending on the analysis)

Analysis of HC cohort

- Should exclude non-HC
- may exclude participants with genetic variants (depending on the analysis)

Analysis of the Prodromal cohorts

- should exclude not prodromal participants (N=9, labeled as 0 in column K)
- the rest of the information is presented to inform the analysis

Please, refer to Summary analytic tab in the Excel spreadsheet Column B LONI nomenclature keys on how to identify participants in each Cohort/ subgroup in LONI

Data presentation

The PPMI analytic dataset is established by the PPMI Clinical Consensus Committee (CC) and will be maintained in Excel format and stored in LONI (until the EDC system develops a platform to record the research tags in and transfer to LONI). The analytic dataset is a living document and will continue to reflect additional PPMI data in the future.

The Excel spreadsheet is organized as follows

1. Summary table (Tab 1) – summary of all participants reviewed by the CC with the Total Numbers of Cohort/ subgroups at **Enrollment and if they changed their classification in the Analytic dataset. Column B includes LONI nomenclature.**
2. Summary Analytic (Tab 2) - summary of the cumulative N of participants in the Cohorts/ subgroups **based on the Analytic dataset**. **Column B includes LONI nomenclature keys on how to identify participants in each Cohort/ subgroup.**
3. Cohorts (Tabs)
    a. Parkinson's disease
    b. Prodromal
    c. Healthy Controls
    d. SWEDD (legacy cohort) – analytic dataset includes ONLY subjects who remain active in the study (N=3)

For each cohort tab, each PPMI participant is listed under the subject ID with columns reflecting their enrollment cohort and subgroup (Columns B-H for PD cohort) and columns reflecting their analytic cohort / subgroup assignment (Columns J-N for PD cohort)

Ongoing collection of clinical, genetic, imaging and biomarker data may result in a change in the enrollment cohort and subgroup. Based on these data, participants may be deleted from their enrollment cohorts or subgroups or may be re-categorized from their enrollment cohort and subgroups to different cohorts or subgroups. Participants with a revised cohort or subgroup are identified with Research Tag - examples below.

Clinical Diagnosis research tag

- There are selected participants recruited into the **PD cohort** who on further review of the data do not have biological/ clinical characteristics consistent with PD and as such were assigned research tag *not-PD*. The objective was not to determine a new diagnosis but to adjudicate with a high degree of certainty re diagnosis of PD. These participants are labeled 0 in Column J (Cohort reassignment) of the attached spreadsheet.
- There are selected participants in the PD genetic cohort who on further review of the data do not have biological/ clinical characteristics consistent with PD and as such were assigned research tag *not-PD but they continue to qualify for genetic prodromal cohort*. These participants are labeled 0 in Column J / and 1 in column K (Cohort reassignment) of the attached spreadsheet.
- There are selected participants recruited into the **HC cohort** who on further review of the data have biological/ clinical characteristics suggestive of degenerative synucleinopathy and as such were assigned

research tag *not-HC*. Such participants should be excluded from the analysis. These participants are labeled 0 in Column H (Cohort reassignment, HC) of the attached spreadsheet

Clinical diagnosis adjudication- was completed by the PPMI Consensus committee that reviewed the compendium of the clinical, imaging and *other available* biomarkers data on all PPMI PD participants where site investigator indicated a change in diagnosis (MSA, PSP, etc).

- In the cases where CC had sufficient confidence in the diagnosis, such diagnosis is posted in the comments section. These participants are labeled 0 in Column J of the attached spreadsheet.
- In some cases, the amount of information was not sufficient to allow diagnosis adjudication and in those cases research tags are not assigned despite site investigator change in diagnosis

Prodromal research tags adjudication

Prodromal research tags include assignment of a new/ revised tags based on the prodromal recruitment subgroups, ie, RBD, hyposmia, genetic LRRK, genetic GBA, genetic SNCA/ or rare mutation.

- (Examples) There may be selected participants recruited into the **prodromal RBD cohort** who based on further genetic testing may carry GBA variant. In that case they will be assigned GBA research tag and labeled 1 in Column M.
- All prodromal participants who meet prespecified criteria for hyposmia (<10% percentile for age/ gender norms) are assigned hyposmia research tag and are labeled 1 in column O. Hyposmia research tags are not assigned for the PD or HC cohorts.
- RBD research tag is assigned only if the participant had PSG confirmed RBD. Considering that PSG data are not available on any participants aside from RBD subgroup, column P is left with a dot (unknown)

Phenoconversion adjudication

CC reviewed all prodromal cases where site investigator determined subjects met criteria for clinically defined neurodegenerative syndrome.

- If the CC adjudicated on the subject meeting criteria for phenoconversion, the subject is labeled as phenoconverted (column Q), columns R/S list diagnosis and visit. Please, note that in case the diagnosis changes with the longitudinal follow up, additional columns will be added with the new diagnosis and visit.
- In case CC did not agree with the determination of phenoconversion, no research tag will be assigned (despite site investigator determination) and the participant will remain listed as prodromal.

Genetic research tag

Genetic data has been and continues to be added during the study to augment participants' genetic characterization. Genetic data were not available at the time of recruitment of the sporadic PD, HC. In addition, participants recruited into either LRRK2 or GBA cohorts originally had focused genotyping related to the target variant. Subsequent genotyping identified participants in these cohorts who carry additional PD relevant genetic variants. The attached spread sheet includes *only* data relevant to GBA and LRRK2, SNCA variants and *only* variants that are considered pathogenic based on the *Recommendations for Determining PPMI Eligibility for Parkinson's Disease Genetics* developed by the genetic core (Methods document posted in LONI). Carriers are labeled 1 for respective variants in the attached spreadsheet. Attached spreadsheet does not include a type of the variant, such data can be found on LONI.

Depending on the nature of the analysis, investigators might choose to exclude participants with pathogenic GBA or LRRK2 variants from the analysis of the sporadic cohort or HC or dual variant carriers from the analysis of the GBA/ LRRK2 cohorts.

Please, note that some of the participants had incomplete dataset (labeled missing) and some variants might be recoded as pathogenic (as additional information becomes available) and as such genetic adjudication might change in the future.

Parkinson's Progression Markers Initiative

SWEDD cohort (legacy)

SWEDD is not a clinical diagnosis but radiological term defined as a Scan without evidence of dopamine deficiency. SWEDDs follow up was limited to 2 years and the cohort follow up has been completed. Full SWEDD dataset is available in LONI. Few (N=3) SWEDDs remain active in PPMI and as such only they are included in the analytic dataset.