Enhancing Neuro Imaging Genetics through Meta-Analysis Consortium (ENIGMA)
Parkinson's Disease Secondary Analysis Proposal

Please complete all fields and return this form by e-mail to: Ysbrand van der Werf, Eva van Heese, Emile d'Angremont at yd.vanderwerf@amsterdamumc.nl; enigma-pd@amsterdamumc.nl;

1. Policy

Members of the ENIGMA Consortium include investigators from different centers around the world who are actively engaged in neuroimaging research and who have contributed results from primary analyses of imaging, genetic data, and/or algorithm development for the purpose of meta-analysis, replication, and/or algorithm testing in a collaborative manner.

Although the data contributed to the ENIGMA consortium consist of group-level summaries and post-estimation statistics rather than raw genotype and phenotype data, there is theoretically a minute risk of determining whether a given individual participated in a study. While the re-identification of samples requires access to the raw genotype data of the target individual and constitutes scientific misconduct, most groups have opted to appoint a gate-keeper approach rather than allowing full public access to the results of their analyses or meta-analyses. Within the ENIGMA-PD working group any consortium member wishing to access the results of specific analyses or meta-analytic results will be asked to complete a short proposal describing why they wish to access the results files from each group, and submit that for review.

All consortium members are encouraged to submit such proposals, to follow up on ideas which the group as a whole cannot pursue, which involve novel analyses, or subsets of the available sites. The ENIGMA-PD working group will screen PD -relevant proposals for scientific interest, and will help enlist members who might be interested in collaborating. Proposals will be discussed on ENIGMA-PD working group calls and emails to encourage the broadest participation.

The proposal will then be posted on an ENIGMA forum page and an email will be sent to all consortium members alerting them to the posting. ENIGMA members will have 14 days from the time of the posting to opt-out of the analysis, ask for clarification, voice concerns or objections and/or give feedback to the proposal. No site data will be shared without the consent of the PI of that site, who may opt to impose specific conditions or limitations on the use of the data; also ENIGMA PIs and members are not required to take part in any proposed project, they can opt out.

If the author of the proposal agrees to the authorship and publication policies of the consortium the access request will be granted to the results files for those groups who have not opted-out of the analysis and a member of the Enigma PD working group or Enigma support group will be assigned as a project liaison. The Enigma support group liaison will be responsible for providing the data and answering any queries relating to the project, and providing the contributing site PIs with updates. If there is no possibility of determining if a particular individual participated in a study (e.g. limited imaging or genetic markers are requested), results from these markers may be sent by the liaison to other sites if available. If genome-wide results are requested from individual groups, the person submitting the proposal may be granted an account on Imaging Genetics Center (IGC) servers or may visit IGC, if desired, to make it easier to complete the analysis. All approved proposals are welcome to use services at IGC. The data can be housed in IGC and will not be transferred or mirrored to other sites.

We request that the 'ENIGMA Consortium' or the specific working group(s), and the liaison person will be listed as co-authors. The ENIGMA Consortium on the byline, or the ENIGMA Working Group on the byline, will reference the PIs of each study, in addition to contributors at their site. In this way the authors contributing data to the consortium will be appropriately acknowledged on any publication.

2. Requestor Information

Date of Submission: 14 Feb 2026

Name project PI: Jean-Baptiste Poline

Email project PI: jean-baptiste.poline@mcgill.ca

Name project runner: Nikhil Bhagwat

Email project runner: nikhil.bhagwat@mcgill.ca

Institution/Affiliation: McGill University

Co-author/collaborator names and e-mail addresses: Tristan Glatard, tristan.glatard@camh.ca

Please list any conflicts of interest: N/A

3. Study proposal

Aim of the project

Proposed title
ParkCore: A PD prediction challenge to assess efficacy of diagnostic and prognostic AI/ML models

Background
Provide an overview of prior research and highlight the challenges or gaps related to the research question.

Global neuroimaging datasets representing diverse populations are essential for assessing efficacy and reliability of hypothesised disease biomarkers and personalized predictive models. The ongoing Artificial Intelligent/Machine-learning (AI/ML) revolution has prompted a surge in diagnostic and prognostic model prototypes, yet their clinical translation remains challenging due to lack of systematic and rigorous validation process.

Competitive prediction challenges leveraging well-defined tasks and common datasets are powerful frameworks for such validation of biomarkers and AI/ML models. Platforms such as Kaggle, over the past decade, have supported AI challenges across various domains which has led to the development of impactful AI models. In brain research, the potential of AI/ML challenges remains underutilized, partly due to difficulties in working with restricted datasets.

This project - part of a larger Brain Health Data Challenge (https://www.bhdc-pdsc.ca/) - proposes to set up a PD prediction challenge leveraging multiple data-hubs (i.e. participating ENIGMA-PD sites) that would evaluate of AI/ML models using local, consistently pre-processed neuroimaging datasets. The challenge provides a separate identically pre-processed dataset (PPMI) for training to the model developers. The trained models are then submitted on a central portal to be tested on unseen, independent datasets. This decentralized setup is designed to ensure local data governance (i.e. test data is NEVER shared with the model developers). The project seeks ENIGMA-PD sites to participate as data

providers and/or model-tester hubs to perform iterative validation of model submissions under this challenge.

Aim
Describe the main aim(s) or research question of the study.
1. Phase-1: Classification of PD vs control using structural imaging-derived-phenotypes (FreeSurfer).

2. Phase-2: Classification of PD vs control using structural (FreeSurfer) and diffusion (QSIPrep) metrics.

3. Phase-3: Prediction of PD symptom progression (motor and non-motor) using structural (FreeSurfer) and diffusion (QSIPrep) metrics.

Note: Sites can choose to participate in any of the phases based on their interest and data availability.

Analyses
Confirmatory analyses (hypothesis-driven)
For each analysis, describe the specific aim and hypotheses derived from literature.
[ ]

Exploratory analyses (data-driven)
For each analysis, describe the specific aim and (if present) hypotheses.

The analysis includes comparisons of classification and prediction performance of submitted AI/ML models across different datasets.

Analysis plan

Data request
Neuroimaging data
Describe the MR images and/or already available imaging-derived phenotypes required for this study. If relevant, describe specifics to the data acquisition (i.e. task, checks, etc).

Phase-1: FreeSurfer 7 outputs (aseg, aprac, samseg)

Phase-2,3: QSIPrep outputs (FA, MD maps aggregated over JHU atlas)

Clinical/demographic data
Describe the required clinical and demographic data to participate in this study.

Required: Age, sex, diagnosis

Nice to have: Education, disease duration, UPDRS, MoCA/MMSE, medication information.

In Phase-3, longitudinal clinical data is needed.

Data processing
Outline the main steps of the data processing workflow, specifying the required software (preferably open-source and standardized pipelines). Provide details on the procedures used for data quality control.

Phase-1

- Image processing with FreeSurfer-7
- Data quality check (QC) with Enigma FSQC protocol.

Note: The Phase-1 data processing and QC should be already done if you participated in the ENGIMA-PD FreeSurfer upgrade workflow.

Phase-2

- Image processing with QSIPrep-1
- Data quality check: TBD

Note: Adoption of the Nipoppy framework to run and track QSIPrep is recommended to simplify and ensure consistency across sites. Technical support by our team can be provided if needed.

Model testing setup

If a site shares processed data (i.e. imaging-derived-phenotypes) with the ENIGMA-PD core site, then no additional setup is required locally. Note that the shared data at ENIGMA-PD core site will not be accessible to the model developers.

If a site prefers to set up a local model tester-hub, then the adoption of Nipoppy and Neurobagel tools will be required to harmonize variables and coordinate decentralized model validation. Our team will provide technical support as needed.

Outcome measures of interest & statistical models
Detail the dependent and independent variables for each research aim. For the statistical models, specify the tests used, any relevant covariates, and applied corrections (e.g., FDR correction, Bonferroni correction, if applicable). Describe the approach for handling potential outliers and exclusions.

For the submitted models, the prediction challenge will assess balanced classification accuracy, false positive and false negative rates across datasets in the Phase 1 and 2. The Phase 3, PD progression tasks will be defined based on availability of longitudinal motor and non-motor symptom scores across sites.

Timeline
Provide the estimated timeline for the project.

- Phase-1
    - June 2026: Deadline to opt-in as a data provider for the challenge
    - July 2026: Submission of data availability (imaging + clinical)
    - Sept 2026: Completion of FS data processing
    - Oct 2026: Internal testing of model validation setup
    - Dec 2026: Phase-1 challenge is live.
- Phase-2
    - Sept 2026: Deadline to opt-in as a data provider for the challenge
    - Oct 2026: Submission of data availability (imaging + clinical)

    - Dec 2026: Completion of FS+QSIPrep data processing
    - Jan 2026: Internal testing of model validation setup
    - March 2027: Phase-2 challenge is live.
- Phase-3
    - Dec 2026: Deadline to opt-in as a data provider for the challenge
    - Jan 2027: Submission of data availability (imaging + clinical)
    - March 2027: Completion of FS+QSIPrep data processing
    - April 2027: Internal testing of model validation setup
    - June 2027: Phase-3 challenge is live.

Other

Anything else you want to mention?

This novel project proposal in ENIGMA is a bit different compared to typical mega analysis projects. The main goal here is not biomarker discovery but rather assessing reliability of AI/ML models across diverse datasets.

The outcomes of this project will therefore be reported in a public leaderboard followed by methods / AI-platform oriented articles. The test datasets and corresponding teams will also be cited in any paper authored by the challenge participants.

We expect the data preparation step to be low-effort and likely to be useful for several other projects. We hope this challenge will serve as a crucial step towards clinical translation of AI/ML models in PD.

Please confirm that you have reviewed the ENIGMA-PD website for potential areas of overlap. If you see a project that may overlap, please list along with any plans for addressing this: [ ]

There is no direct overlap with on-going projects. The Multiple Symptom Progression Modelling project possibly uses similar imaging features but focuses on disease mechanisms, whereas the Phase-3 of the proposed project aims at assessing individual-level predictions by AI/ ML models.

References

1. Brain Health Data Challenge: https://www.bhdc-pdsc.ca/
2. Nipoppy: https://nipoppy.readthedocs.io/en/latest/
3. Neurobagel: https://neurobagel.org/