

Publishable Summary

This deliverable reports on the integration of multiple data sources for the purposes of the AMYPAD PNHS final analyses and access to internal and external researchers. This activity is crucial for the project as the PNHS study enrolls subjects across multiple Parent Cohorts. Therefore, there are two steps to integration: first, the imaging data prospectively collected in AMYPAD PNHS needs to be collated with the corresponding non-imaging variables collected by the Parent Cohorts; second, the complete data-sets from each Parent Cohort need to be integrated into a final data-set which can be used for the main analyses in the study. This second step requires an additional phase of harmonization, as the variables across multiple cohorts are equivalent (record the same information), but not identical.

To address these needs, three teams have been set-up. The first, a Harmonization Team, was put together to dedicate efforts to the harmonization of non-imaging variables from all Parent Cohorts. This team meets once a week to define a harmonization strategy per variable available across cohorts, arriving at a consensus on the target variables (e.g. weight), definitions (e.g. measured weight), and format (e.g. weight in kg). At this moment, only the proposal for the cognitive outcomes is final (main objective of the study), while other variables (e.g. anthropometric measures, medical history, socio-demographic status, and lifestyle and risk factors) are still being defined as some information from some Parent Cohorts is still being retrieved. Nonetheless, initial proposals are in place. The second, an Infrastructure Team, was also set-up with the addition of Aridhia as a subcontractor to the project. This team meets twice a week, and is responsible for establishing a two-layer process for the integration of the data. First, several Castor EDCs are constructed to host the original data coming from the Parent Cohorts. Then, these are transferred to the Aridhia Workspaces, where they are processed and integrated into an overarching data-set across all cohorts. Finally, the third team, Imaging Harmonization, has set up an SOP for the collection of phantom scans across sites with Parent Cohorts that have historical (pre-AMYPAD) scans to be collected and quantified. At this stage, all imaging site leads have been contacted and, along the upcoming year, one person from AMYPAD will be visiting the imaging sites to perform the phantom acquisitions in the most comparable way. Then, this information will be processed in a pre-established way to render comparable quantitative measurements of amyloid load as measured by PET across sites and cohorts.

Overall, the different teams are progressing well with setting up 1) non-imaging and 2) imaging data harmonization procedures as well as 3) the infrastructure for data integration. The first Parent Cohort has already transferred data (EMIF-AD) to AMYPAD, and we have access to EPAD LCS as well. Structures for receiving data from ALFA+ and FACEHBI are also already in place, and the first transfers will take place in Q4 2020. It is also expected that the final strategy for harmonization of non-imaging variables will be final in Q1 2021, allowing all

data-sets to come in and AMYPAD PNHS to begin a complete interim analyses across all Parent Cohorts.

For more information: info@amypad.org

Acknowledgement: This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115952. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.



innovative
medicines
initiative

