ADNIMERGE2 R Data Package

Authors: Micheal C. Donohue, PhD[1]; Kedir Hussen, MSc[1]

Institution: [1]USC Epstein Family Alzheimer's Therapeutics Research Institute, Keck School of Medicine, University of Southern California, 9860 Mira Rim, San Diego, CA 92121.

1. Introduction

ADNIMERGE2 is a new version of the existing ADNIMERGE R data package developed for the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. All study datasets are included in the new package version as shared on LONI.

2. Key Features of the New R Package

The latest version of the R package introduces several enhancements and functionalities, including:

- *Integration of Current Study Data*: Incorporates data from ongoing *ADNI4* study phase as shared on LONI unlike the previous version that was limited to ADNI1-3 study phases.

- *Modified PACC Scoring Methodology*: Provides a modified version of the Preclinical Alzheimer Cognitive Composite (PACC) score dataset along with a function to compute the score.

- *Derived/Analysis Datasets:* Includes newly created datasets that utilize the 'pharmaverse' framework for creating derived data. Please see section 3.2 for more details.

- *Articles/Vignettes*: Provides detailed guidance on how to build an R package from raw csv files, computing PACC scores, creating derived/analysis datasets, and simple enrollment and longitudinal summaries. Please refer to section 4 for more details.

- *Utility Functions*: Integrates some common R functions that can be used for data preparation purpose, particularly for ADNI study data. For examples, functions that will be used to flag enrolled participants or participants who had fatal event.

Note that the "Key ADNI Table Merged Data (adnimerge)" will no longer be provided in the new version of the ADNI R package.

3. Dataset List

Table 1 Dataset type

Raw Dataset	Defined all the study data that are available/downloaded from LONI
Derived Dataset	Defined data that were created during the ADNIMERGE2 R package build and that are included in the package. This includes PACC score data, and other derived/analysis datasets

3.1. Raw Datasets

Table 2 shows list of raw datasets that are available in *ADNIMERGE2* R package. The complete list of available raw datasets can be found in the Reference section of the ADNIMERGE2 website under "*ADNI Raw Dataset*".

Table 2 List of raw datasets

ADSP Phenotype Harmonization Consortium	Enrollment
Biospecimen Result Table[1]	Imaging Acquisition and Result Table
Cognitive/Functional Assessment	Medical/Health & Adverse Events Log History
Data Dictionary	Neuropathology Table
Diagnostics Summary Table	Remotely Collected Data (ADNI Online)

[1] *Only selected datasets*

Snapshot of Raw Data Preparation Steps:

- *Raw data files were downloaded from LONI on the same date.*

- *All raw datasets were named in **uppercase** in version, unlike the previous version where dataset named in lowercase.*

- *Two variables were created in the raw data to identify the data collection and original study phase. Both variables were also presented in previous version of the package.*
  - **COLPROT**: S*tudy protocol of data collection is a renamed version of the original PHASE, Phase, or ProtocolD variable. Only included in the raw data if the actual file contains one of these variables.*
  - **ORIGPROT:** Original study protocol either *ADNI1*, *ADNI-GO*, *ADNI2*, *ADNI3* or *ADNI4* based on participant participation history in ADNI study.

- *Values of **-4** or **-1** were converted into missing values per R software standard.*

- Numerically coded values were mapped according to the data dictionary file, and variables that were successfully mapped are flagged in the corresponding data documentations.

- No additional data wrangling was performed on the raw data.

3.2. Derived Datasets

As part of the R package development, few derived datasets are also included in the package. The main rational of including these datasets is to illustrate the '*pharmaverse'* framework. Please refer to the following *ADNMERGE2 Derived Data* and *ADNIMERGE Analysis Data* articles/vignettes for detailed description on how those datasets were generated. The list of those datasets can be also found in Reference section of the ADNIMERGE2 website under "*ADNI Derived Dataset*".

Furthermore, a modified version PACC score data is also included in the package and the corresponding detailed method can be found at Articles section of the ADNIMERGE2 website under "*PACC Score Modified Version*". The PACC score data will be also available soon on LONI as a separate dataset in csv file format.

NOTE:

- Note that these derived datasets may not fully compile with the CDISC standardization.
- It is highly recommended to review the package website for detailed information on how derived/analysis datasets were generated before using them in analysis.

4. Articles/Vignettes

Table 3 shows the complete list of articles that are included in the latest ADNIMERGE2 R package. All those articles can be accessed using '*browseVignettes("ADNIMERGE")*' code in the R software after installing the study package locally.

Table 3 ADNIMERGE2 articles or vignettes

Vignettes	Description
ADNIMERGE2	An overview/Get Started page about the study package
ADNIMERGE2-Package-Building	To demonstrate how to build an R package from raw csv files
ADNIMERGE2-PACC-SCORE	Provides detailed methods on how to compute a modified version of PACC score in ADNI study
ADNIMERGE2-Derived-Data *ADNIMERGE2-Analysis-Meta-Specs* *ADNIMERGE2-Analysis-Data*	Provides detailed methods on how derived/analysis datasets in this package were generated
ADNI-Enrollment	To demonstrate how to use the study package to create simple enrollment summaries across study phase, and baseline diagnostics status
ADNI-Longitudinal	To illustrate how to use the study package to create summary of longitudinal cognitive assessment

5. Website

Package Website URL: https://atri-biostats.github.io/ADNIMERGE2

GitHub URL: https://github.com/atri-biostats/ADNIMERGE2/

6. References

Donohue MC, Hussen K, Langford O, Gallardo R, Jimenez-Maggiora G, Aisen PS. Alzheimer's Clinical Research Data via R Packages: the alzverse. *arXiv preprint* arXiv:2510.02318. 2025 Sep 18.