More about data

We provide some fundamental information about using the ADNI data files. Topics of common confusion when first starting out working with ADNI data or trying to navigate through changes in data acquisition across the phases of ADNI are also presented.

- Data Organization
  - Data within a file are organized with each visit in a separate row (the 'long' format of data rather than the 'wide' format)
  - RID identifies rows belonging to the same individual.
- Notes about data
  - Missing data codes: Most use '-4' as missing, although a few will say '-1'. In some files, blank cells are also used for missing data. Check the data dictionary (DATADIC.csv) for missing value codes.
  - Common variables for linking files
    - RID (unique participant identifier)
    - VISCODE (visit code) - in ADNIGO/2/3, these codes are not meaningful to the user (tracking codes)
    - VISCODE2 (visit code) - more meaningful month visit codes (i.e. sc, bl, m06, m12, etc.)
    - EXAMDATE (date of assessment - not available in all files and some information collected at a 'visit' may happen over several days; for example, imaging may be done on a different day)
  - Two 'initial visits': sc and bl
    - Some information is collected at screen while other information is collected at baseline.
    - Important to keep in mind when merging data for the initial visit, since these span over two actual visits.
  - VISCODE='f' means the subject failed screening (ADNI1)
- Diagnosis
  - 'Initial diagnosis' - actually two different ones
    - Diagnosis at the screen visit (VISCODE2='sc')
      - When randomization is assigned
      - Determines visitation schedule, which is diagnosis specific
      - Available in the arm table (ARM.csv)
    - Diagnosis at baseline visit (VISCODE2='bl')
      - Based on additional information acquired at the baseline visit to obtain diagnosis
      - Diagnosis recommended by the Clinical Core for use as initial diagnosis
      - Available in the diagnosis summary table (DXSUM_PDXCONV_ADNIALL.csv)
      - Note that in ADNIGO/2, if you want to identify SMC and EMCI, you need to use the information in the arm table and verify that the diagnosis was stable at VISCODE2='bl' in the diagnostic summary table (SMC and EMCI are not diagnostic categories used after the screen visit).
  - Current diagnosis variable names change by the phase in the diagnosis summary table.
    - ADNI1: DXCURREN 1=NL; 2=MCI; 3=AD
    - ADNIGO/2: DXCHANGE 1=Stable: NL to NL; 2=Stable: MCI to MCI; 3=Stable: Dementia to Dementia; 4=Conversion: NL to MCI; 5=Conversion: MCI to Dementia; 6=Conversion: NL to Dementia; 7=Reversion: MCI to NL; 8=Reversion: Dementia to MCI; 9=Reversion: Dementia to NL
    - ADNI3: DIAGNOSIS 1=CN; 2=MCI; 3=Dementia

More about data

- About EXAMDATE in Clinical data files
  - Clinical data acquired in ADNIGO/2 do not include 'EXAMDATE' (the date of the exam), although this information was included for ADNI1 visits.
  - The variable 'USERDATE' is the data entry (or modification) date, and may be very different from EXAMDATE.
  - Use the variable 'EXAMDATE' in the registry table (REGISTRY.csv) as the date of exam for all clinical data (merge by RID and VISCODE or VISCODE2)
- ADNIMERGE file
  - For many of the commonly used data, a linked file is available for download (ADNIMERGE.csv).
  - Includes demographics, neuropsychological testing scores, MRI and PET summaries, CSF measures
  - Also includes both a baseline diagnosis and diagnosis at each visit