Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

ADNI Data Training Part 2

ADNI Biostatistics Core Team

UC Davis School of Medicine
Department of Public Health Sciences

August 1, 2013

ADNI Alzheimer's Disease Neuroimaging Initiative

UCDAVIS
UNIVERSITY OF CALIFORNIA

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Outline

Today's Presentation Outline

- Data overview

- Commonly used data (tables)

- Helpful tips about working with data

- Cross-validation

- IDA image data

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

About ADNI Data

Data organization

- RID: Participant roster ID.
- In general, data is 'long' format
  - ✓ One row per each visit (VISCODE)
  - ✓ Multiple rows belonging to same subject.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

About ADNI Data

Data organization

- Some exceptions in image data.

  ✓ ADNI1:
     - Some MRI files have both 1.5T and 3T scans (FLDSTRENG or MRFIELD).
     - Stroke, WMH file has one row per stroke (STROKESUM_V2.csv and MRI_INFARCTS.csv)
     - FDG-PET: UC Berkeley has 5 different regions, one per row, for each visit.
  ✓ ADNIGO2:
     - MRI files may include both accelerated and non-accelerated scans.
     - MAYOADIRL_MRI_MCH.csv has one row per MR finding.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

About ADNI Data

Visit code

- VISCODE
  - ✓ ADNI2 uses different convention.
  - ✓ ADNI2_VISITID table links VISCODE to VISCODE2.
- VISCODE2
  - ✓ All phases use same convention(i.e. sc, bl, m06)
  - ✓ Not all data contains VISCODE2.

Phase

- PHASE or COLPROT: ADNI1, GO or 2.
- Some files do not contain this variable. (i.e. PET data, MRI data)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

About ADNI Data

Date variables

- EXAMDATE: date of the exam.
  - ✓ Most clinical data files do not have EXAMDATE for ADNI2.
  - ✓ Please extract EXAMDATE from REGISTRY table using RID and VISCODE.
- USERDATE: date of data entry (maybe very different from EXAMDATE).

Missing values

- Missing data is coded with -1 and -4.
  - ✓ -4: 'passively' missing or not applicable.
  - ✓ -1: confirmed missing at point of data entry

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

LONI: Download Data

You can download data from the **Download Study Data** page.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Assessments

Assessments > Diagnosis

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Assessments

Assessments > Diagnosis
Diagnostic Summary (DXSUM_PDXCONV_ADNIALL.csv)

- Diagnosis by each visit code.
- ADNI1 and ADNIGO2 use different variables for diagnosis.
    - ✓ ADNI1: DXCURREN, DXCONV, DXREV, DXCONTYP.
    - ✓ ADNIGO2: DXCHANGE.
- If you use ADNIMERGE package, DXCHANGE is the only diagnosis variable.
- We will discuss more in the Working with Data section.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Assessments

Assessments > Neuropsychological

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Assessments
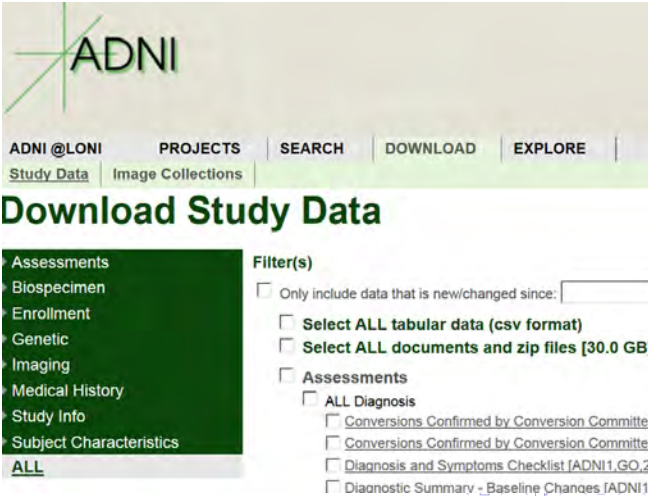
Assessments > Neuropsychological
**ADAS Sub-Scores and Total Scores[ADNI1]
(ADASSCORES.csv)**

- ADAS-cog sub and total scores in ADNI1.
- Key variables:
  - ✓ TOTAL11: 11 items score
  - ✓ TOTALMOD: 13 items score

**Alzheimer's Disease Assessment Scale (ADAS)[ADNIGO,2]
(ADAS_ADNIGO2.csv)**

- ADAS-cog total scores in ADNIGO/2.
- Key variables:
  - ✓ TOTSCORE: 11 items score
  - ✓ TOTAL13: 13 items score

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Assessments

Assessments > Neuropsychological
Clinical Dementia Rating Scale(CDR) (CDR.csv)

- 6 domains and global scores are available.

  ✓ CDMEMORY: memory , CDORIENT: orientation , CDJUDGE:
    judgement & problem solving , CDCOMMUN: community
    affairs , CDHOME: home and hobbies , CDCARE: personal
    care , CDGLOBAL: Global CDR

SAS example code to create CDRsum score

```
array cdr {*}cdmemory cdorient cdjudge cdcommun cdhome cdcare ;
do j=1 to dim(cdr) ;
  if cdr{j} in ( -1, -4 ) then cdr{j} = . ;
end;
if nmiss(of cdr{*})=0 then cdrsum = sum(of cdr{*}) ;
label cdrsum = "Clinical Dementia Rating Sum " ;
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Assessments

Assessments > Neuropsychological
Everyday Cognition-Participant Self Report[ADNIGO,2] (ECOGPT.csv)

- Create each domain score by taking average (at least half of the items are not missing for each domain)
    - ✓ memory 8 items (memory1-memory8)
    - ✓ language 9 items (lang1-lang9)
    - ✓ visuo-spatial 7 items (visspat1-visspat4, visspat6-visspat8 : visspat5 is a duplicated field (see DATADIC.csv))
    - ✓ planning 5 items (plan1-plan5)
    - ✓ organization 6 items (organ1-organ6)
    - ✓ divided attention 4 items (divatt1-divatt4)
    - ✓ total score 39 items

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Assessments

Assessments > Neuropsychological
Functional Activities Questionnaires (FAQ.csv)

- FAQ item and total score.
  - ✓ FAQTOTAL: FAQ total score.

Mini-Mental State Examination (MMSE.csv)

- MMSE item and total score.
  - ✓ MMSCORE: MMSE total score.

Neuropsychological Battery (NEUROBAT.csv)

- All remaining neuropsychological test scores.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Biospecimen

Biospecimen > Biospecimen Results

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Biospecimen

Biospecimen > Biospecimen Results
ApoE-Results (APOERES.csv)

- ApoE Genotyping Results
  - ✓ APGEN1: Genetype - Allele 1 (2,3, or 4)
  - ✓ APGEN2: Genetype - Allele 2 (2,3, or 4)
- PHASE: it has either 'ADNI1' or 'ADNIGO2'
- For ADNIGO2, it has missing values on VISCODE, SITEID, and APTESTDT(examdate) as of June 2013.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Biospecimen

Biospecimen > Biospecimen Results
UPENN CSF Data (ABETA, TAU, PTAU)

- √ 4 sets of data from ADNI1, one dataset from ADNIGO/2, and one dataset from ADNI1/GO/2.

- √ Each data contains results of each analytical runs, so two data could have different values for the same subject at the same visit time..

- √ Researchers should only use values within the same file. (see LONI website Data FAQs)

- √ UPENNBIOMK5.csv file will be updated soon (from LONI Biomarker Core News as of July 23, 2013): Biomarker Core is re-anchoring the 2012 results using most recent data.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Biospecimen

Biospecimen > Biospecimen Results
UPENN-Biomarker Data (UPENNBIOMK.csv):

- baseline abeta, tau, ptau for 415 ADNI1 subjects (+1 screening failed subject: RID=975)

UPENN-Longitudinal Data (UPENNBIOMK2.csv):

- baseline&m12 abeta and tau for 417 ADNI1 subjects (92 subjects have missing value on m12)

UPENN-Longitudinal Data(3yr) (UPENNBIOMK3.csv):

- baseline(n=103), m12(n=101), m24(n=87)&m36(n=23) abeta, tau, ptau for ADNI1 subjects.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Biospecimen

UPENN-Longitudinal Data(4yr) (UPENNBIOMK4.csv):

- baseline(n=141), m12(n=138), m24(n=102), m36(n=78)&m48(n=33) abeta, tau, ptau for ADNI1 subjects (+1 screen failed subjects have bl,m12,m36: RID=975).

UPENN-CSF Biomarkers[ADNIGO2](UPENNBIOMK5.csv):

- baseline abeta, tau, and ptau for 117 ADNIGO subjects and 271 ADNI2 subjects (+1 screen failed subject: RID=4124)
- This file will be available soon (as of Jul 23, 2013).

Second batch analysis of CSF (UPENNBIOMK6.csv):

- longitudinal abeta, tau, ptau for 82 ADNI1 subjects (up to m84: 4 subjects), bl and m24 for 32 ADNIGO subjects(n=5 have missing bl), and baseline for 309 ADNI2 subjects.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Biospecimen

Biospecimen > Lab Collection Procedures
Laboratory Data (LABDATA.csv)

- Screening clinical lab results (i.e. urine, chemistry panel).
- Data contains some character coding (i.e. SCC09: No specimen received), and they can be treated as missing data.
- Currently, ADNI1/GO lab results are available on LONI. (ADNIMERGE package contains ADNI2 lab results also.)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Enrollment

Enrollment > Enrollment

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Enrollment

Enrollment > Enrollment

ADNI2 Visit Codes Lookup[ADNI2] (ADNI2_VISITID.csv)

- Visit code assignment for each ADNI2 subjects.

ARM[ADNI1, GO, 2] (ARM.csv)

- Arm assignment
- EMCI and SMC information can be obtained.
    - ✓ ARM: 1=NL(ADNI1 1.5T only), 2=LMCI(ADNI1 1.5T only) ,
      3=AD(ADNI1 1.5T only) , 4=NL(ADNI1 PET+1.5T) ,
      5=LMCI(ADNI1 PET+1.5T) , 6=AD(ADNI1 PET+1.5T) ,
      7=NL(ADNI1 3T+1.5T) , 8=LMCI(ADNI1 3T+1.5T) ,
      9=AD(ADNI1 3T+1.5T) , 10=EMCI , 11=SMC

Early Discontinuation and Withdrawal(TREATDIS.csv)

- list of subjects who discontinued from the study.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Enrollment

Enrollment > Enrollment

Registry[ADNI1,GO,2](REGISTRY.csv)

- Contains important key variables.
  - ✓ EXAMDATE: date of assessment (clinical data for ADNIGO/2 do not include this field, so you need extract EXAMDATE from this table)
  - ✓ RGCONDCT: whether this visit was conducted (ADNI1)
  - ✓ PTSTATUS: whether active or discontinued from follow up.
  - ✓ RGSTATUS: whether screening visit was performed

Roster[ADNI1,GO,2](ROSTER.csv)

- List for RID and PTID(ADNI subject ID: form of 123_S_5678)

Visits[ADNI1,GO,2](VISITS.csv)

- Dictionary of VISCODE. (ADNI2 use different convention.)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Genetic

Genetic > Genetic Results

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Imaging

Imaging > MR Imaging Analysis

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Imaging

Imaging > MR Imaging Analysis

- Data comes with its data dictionary and method paper.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Imaging

Imaging > PET Imaging Analysis

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Study Info

Study Info > Data & Database

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Study Info

Study Info > Data & Database

**ADNI 1.5T MRI Standardized Lists
(ADNI_1.5T_MRI_Standardized_Lists.zip)**

**ADNI 3T MRI Standardized Lists
(ADNI_3T_MRI_Standardized_Lists.zip)**

- Standardized analysis sets of volumetric scans from ADNI1.

Data Dictionary[ADNI1,GO,2] (DATADIC.csv)

- Data dictionary of most of data on LONI.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Study Info

Study Info > Data & Database

Key ADNI tables merged into one table (ADNIMERGE.csv)

- contains some of the key variables in one table.

Merged ADNI1/GO/2 Packages

- ADNI Merge packages for R, SAS, SPSS, and Stata. (we will talk more about ADNIMERGE packages later)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Subject Characteristics

Subject Characteristics

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Commonly Used Data: Subject Characteristics

Subject Characteristics

Family History Questionnaire (FHQ.csv)

- information of parents and if they have siblings.

- yes=1/no=0/don't know=2 if their mother or father have dementia or having AD.

Family History Questionnaire Subtable (RECFHQ.csv)

- information of siblings (if they have siblings in FHQ.csv).

- yes=1/no=0/don't know=2 if they have dementia or having AD (one row per each sibling).

Subject Demographics (PTDEMOG.csv)

- Demographic information at screening (for each phase).

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

ADNIMERGE Packages

What is the ADNIMERGE package?

- This loads all ADNI data (except genetic data).
- R, SAS, Stata and SPSS versions are available.
- Mike Donohue from UC San Diego wrote R code to store R dataframes in SAS, Stata, and SPSS.
- It includes 'adnimerge' data which contains commonly used variables. This single csv file is also available to download. (i.e. demographic, clinical exam, MRI and PET variables)
- Labels & formatting have been incorporated in R, SAS, Stata.
- ADNIMERGE packages are updated daily.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

ADNIMERGE: R users

Download `ADNIMERGE_0.0.1.tar.gz`

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
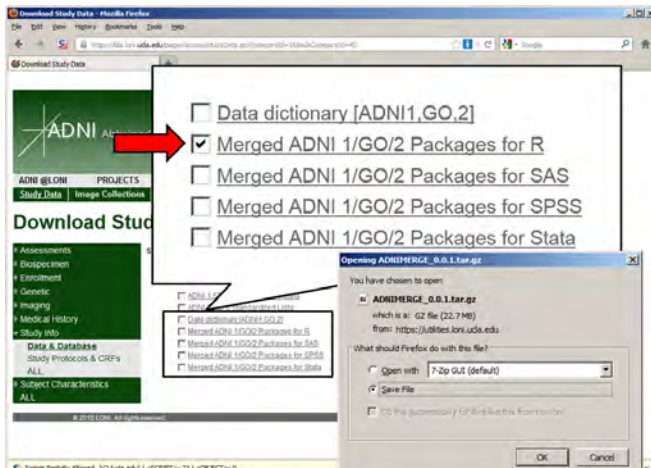
ADNIMERGE: R users

Open R.

Install `Hmisc` package if it is not installed already.

```
install.packages("Hmisc")
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
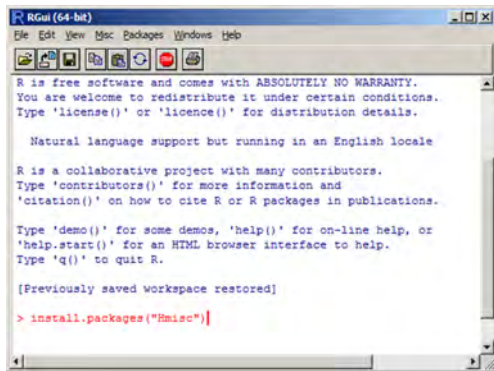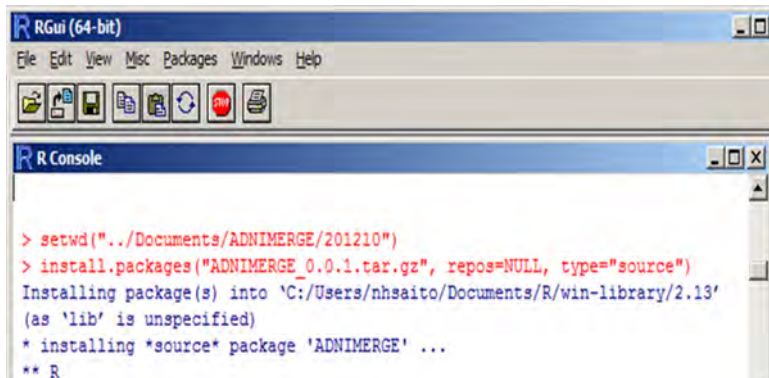
ADNIMERGE: R users

Install the `ADNIMERGE` package which you have downloaded.

```
install.packages("../your path/ADNIMERGE_0.0.1.tar.gz",
repos=NULL, type="source")
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

ADNIMERGE: R users

To load the ADNIMERGE package.

library(ADNIMERGE)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

ADNIMERGE: R users

- To see the documentation, `help(package="ADNIMERGE")`.
- `ADNIMERGE` package loads all ADNI data, so you can start working with individual tables.
    - ✓ One of the loaded items is the `adnimerge` dataframe which contains commonly used variables. (i.e. demographics, clinical exam, MRI and PET)
- For more information, from Windows Explorer, you can open/extract `ADNIMERGE_0.0.1.tar.gz` file using `7-zip`. (7-zip is an open source file archiver designed for Windows.)
    - ✓ You see `data`, `inst`, `man`, `R` folders; `inst` folder contains useful examples.
    - ✓ `README` file contains instructions for how to use the package.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

ADNIMERGE: SAS users

Download `ADNIMERGE_SAS.zip`

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
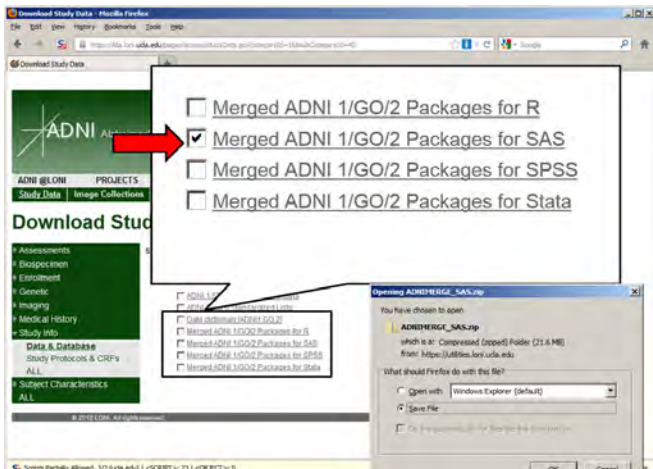
ADNIMERGE: SAS users

Extract ADNIMERGE_SAS.zip file.

Under ADNIMERGE_SAS folder, you will find code folder and data folder.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
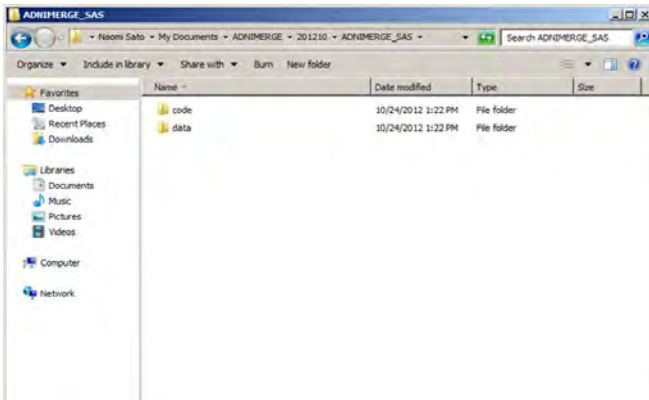
ADNIMERGE: SAS users

In the `code` folder, there are more than 120 SAS programs.

For example, we can open `adnimerge.sas`.

- The simplest method is double-click on `adnimerge.sas`.
- This initiates SAS and opens the programming file.
- This also sets the working directory under the `code` folder.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
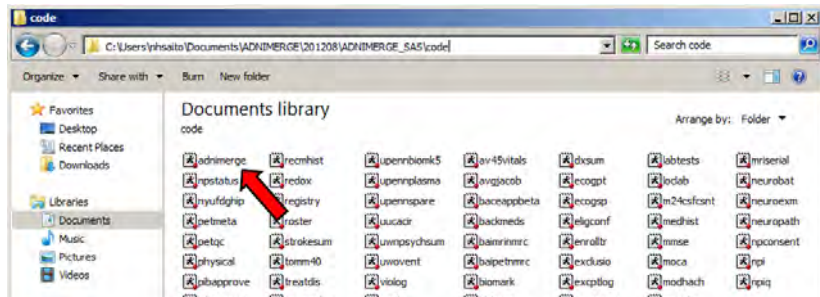
ADNIMERGE: SAS users

You can Run the program.

- Because we set the working directory under the code folder, the program can call the data file correctly.
- SAS will create data called adnimerge under SAS's Work library.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
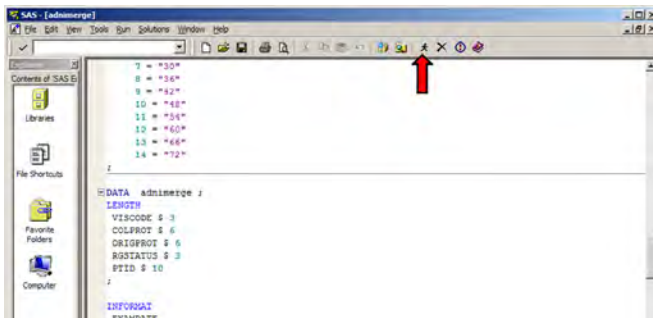
ADNIMERGE: Other users

SPSS package is similar to SAS packages.

Stata users:

- After you extract zip file, you will see more than 120 Stata data(.dta).
- The simplest method: double click the data you want to open, and it initiates Stata and opens data.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
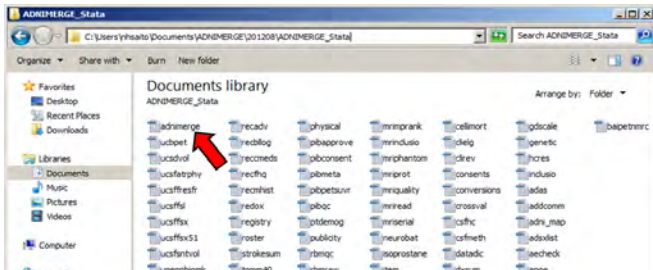
Useful Tips for Using ADNI Data

- Common variables for linking tables
  - ✓ RID
  - ✓ VISCODE (or VISCODE2)
  - ✓ EXAMDATE (date of assessment)
- VISCODE='f' means the subject failed screening (ADNI1).
- RID can tell which PHASE enrolled subject initially.
  - ✓ $RID < 2000$: ADNI1 subjects
  - ✓ $2000 \leq RID < 4000$: ADNIGO subjects
  - ✓ $RID \geq 4000$: ADNI2 subjects
- About EXAMDATE
  - ✓ Clinical data in ADNIGO2 do not include EXAMDATE.
  - ✓ Use REGISTRY.csv to extract EXAMDATE for clinical data.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Useful Tips for Using ADNI Data

Before analyzing data, taking following steps may be helpful.

1. DXSUM data: Since ADNI1 and ADNIGO2 use different variables for diagnosis, assign the diagnosis variable:DXCHANGE for ADNI1.

2. Merge DXSUM and ARM tables and assign baseline diagnosis (including EMCI/SMC).

3. Merge REGISTRY and DXSUM/ARM tables to have EXAMDATE for all visits.

Note: We can identify EMCI or SMC at baseline (or screening) time only. The variable: DXCHANGE is available for follow-up visits, and it tells us either Normal, MCI, or AD.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Useful Tips for Using ADNI Data

1: Use ADNI1 diagnosis variables to assign DXCHANGE.
(DXSUM_PDXCONV_ADNIALL)

- ADNI1 diagnosis variables are following:

    - ✓ DXCURREN: 1=NL, 2=MCI, 3=AD.
    - ✓ DXCONV: 0=No, 1=Yes-Conversion, 2=Yes-Reversion.
    - ✓ DXREV: 1=MCI to Normal, 2=AD to MCI, 3=AD to Normal.
    - ✓ DXCONTYP: 1=Normal to MCI, 2=Normal to AD, 3=MCI to AD.

- ADNIGO 2 diagnosis variable:

    - ✓ DXCHANGE: 1=Stable:NL to NL, 2=Stable:MCI to MCI, 3=Stable:AD to AD, 4=Conv:NL to MCI, 5=Conv:MCI to AD, 6=Conv:NL to AD, 7=Rev:MCI to NL, 8=Rev:AD to MCI, 9=Rev:AD to NL.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Example code

R example code to assign DXCHANGE for ADNI1

```
dxsum = read.csv("DXSUM_PDXCONV_ADNIALL.csv")
attach(dxsum)
dxsum$DXCHANGE[DXCONV==0 & DXCURREN==1] = 1
dxsum$DXCHANGE[DXCONV==0 & DXCURREN==2] = 2
dxsum$DXCHANGE[DXCONV==0 & DXCURREN==3] = 3
dxsum$DXCHANGE[DXCONV==1 & DXCONTYP==1] = 4
dxsum$DXCHANGE[DXCONV==1 & DXCONTYP==3] = 5
dxsum$DXCHANGE[DXCONV==1 & DXCONTYP==2] = 6
dxsum$DXCHANGE[DXCONV==2 & DXREV==1] = 7
dxsum$DXCHANGE[DXCONV==2 & DXREV==2] = 8
dxsum$DXCHANGE[DXCONV==2 & DXREV==3] = 9
detach(dxsum)
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Useful Tips for Using ADNI Data

- 2: Assign baseline diagnosis: including EMCI and SMC.

  - Merge DXSUM_PDXCONV_ADNIALL and ARM tables using RID and PHASE.

  - Keep key variables.

    - ✓ DXSUM: RID, PHASE, VISCODE, VISCODE2, DXCHANGE.
    - ✓ ARM: RID, PHASE, ARM, ENROLLED.

  - Use baseline DXCHANGE and ARM to assign baselineDX variable. (see code in next slides)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Example code

SAS example code to merge ARM and DXSUM

```
* sort data first ;
PROC SORT DATA=dxsum;
  BY rid phase;
PROC SORT DATA=arm;
  BY rid phase;
RUN;

* merge data;
DATA dxarm;
  MERGE dxsum(KEEP=rid phase viscode viscode2 dxchange)
    dxarm(KEEP=rid phase arm enrolled);
  BY rid phase;
RUN;
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Example code

SAS example code to create baseline diagnosis

```
DATA baseData;
  SET dxarm(WHERE= (viscode2='bl' and enrolled in(1,2,3)));
  * pls format them as 1:Normal,2:SMC,3:EMCI,4:LMCI,5:AD,;
  IF dxchange in(1,7,9) & arm NE 11 THEN baselineDx=1;
  ELSE IF dxchange in(1,7,9) & arm=11 THEN baselineDx=2;
  ELSE IF dxchange in(2,4,8) & arm=10 THEN baselineDx=3;
  ELSE IF dxchange in(2,4,8) & arm NE 10 THEN baselineDx=4;
  ELSE IF dxchange in(3,5,6) THEN baselineDx=5;
RUN;

* merge baseline diagnosis data and dxarm;
DATA dxarm;
  MERGE dxarm baseData(KEEP = rid baselineDx);   BY rid;
RUN;
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Example code

R example code to merge ARM and DXSUM

```
# readin csv files
arm <- read.csv("ARM.csv")

# identify variable to keep for merged data
armVars <- c("RID","Phase","ARM","ENROLLED")
dxsumVars <- c("RID","Phase","VISCODE", "VISCODE2",
"DXCHANGE")

# merge data
dxarm <- merge(subset(dxsum, select=dxsumVars), subset(arm,
select=armVars), by=c("RID", "Phase"))
# baseline data
baseData <- dxarm[dxarm$VISCODE2=='bl' & dxarm$ENROLLED
%in% c(1,2,3),]
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Example code

R example code to assign baseline diagnosis

```
# assign baseline diagnosis

attach(baseData)
baseData$baselineDx[(DXCHANGE %in% c(1,7,9)) & ARM != 11 ] = 1
baseData$baselineDx[(DXCHANGE %in% c(1,7,9)) & ARM == 11 ] = 2
baseData$baselineDx[(DXCHANGE %in% c(2,4,8)) & ARM == 10 ] = 3
baseData$baselineDx[(DXCHANGE %in% c(2,4,8)) & ARM != 10 ] = 4
baseData$baselineDx[(DXCHANGE %in% c(3,5,6)] = 5
detach(baseData)
# merge baseline diagnosis
baseVars <- c("RID","baselineDx")
dxarm <- merge( dxarm, subset(baseData, select=baseVars),
by=c("RID"))
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Useful Tips for Using ADNI Data

3: Keep EXAMDATE.

- Merge REGISTRY table and merged ARM/DXSUM table using RID, PHASE, and VISCODE.
- Keep key variables
    - ✓ REGISTRY: PHASE, RID, VISCODE, VISCODE2, EXAMDATE, PTSTATUS, RGCONDCT, RGSTATUS, VISTYPE.

- EXAMDATE in REGISTRY table is necessary when you merge with other clinical data for ADNIGO/2.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Example code

SAS example code to merge ARM/DXSUM and REGISTRY

```
* sort data first ;
PROC SORT DATA=dxarm;
  BY rid phase viscode;
PROC SORT DATA=registry;
  BY rid phase viscode;
RUN;

* merge data;
DATA dxarm_reg;
  MERGE registry(KEEP=rid phase viscode viscode2 examdate ptstatus
  rgcondct rgstatus vistype) dxarm;
  BY rid phase viscode;
RUN;
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Example code

R example code to merge ARM/DXSUM and REGISTRY

```
# readin csv files
registry <- read.csv("REGISTRY.csv")

# identify variable to keep for merged data
regVars <-c("RID", "Phase", "VISCODE", "VISCODE2",
"EXAMDATE", "PTSTATUS", "RGCONDCT", "RGSTATUS",
"VISTYPE")

# merge data
dxarm_reg <- merge(dxarm, subset(registry, select=regVars),
by=c("RID", "Phase", "VISCODE"))
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects in ADNI1

ADNI1 Study

- 822 subjects passed screening but only 819 had baseline observation.

  - Identified by PHASE='ADNI1', VISCODE='bl', and RGCONDCT=1 in merged REGISTRY/DXSUM/ARM table.

- Randomized to one of 3 arms and baseline diagnosis.

ARM	NL	MCI	AD	Total
1.5T Only	59	94	44	197
PET + 1.5T	107	210	102	419
3T + 1.5T	63	93	47	203
Total	229	397	193	819

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects in ADNI1

ADNI1 Study

- Lumbar puncture was not mandatory, and we have 415 subjects with baseline CSF in UPENNBIOMK.csv

	NL	MCI	AD	Total
Baseline CSF	114	199	102	415

- All 819 subjects have ApoE Genotyping results (APOERES.csv)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects in ADNIGO

ADNIGO Study

- New 128 subjects had baseline observation (Note: 127 EMCI, 1 subject reversion to Normal at baseline).

    - Identified by PHASE='ADNIGO', VISCODE='bl', and PTSTATUS=1(Active) in merged REGISTRY/DXSUM/ARM table(N=131). However, 3 subjects (RID:2071, 2314, 2351) were noted as early withdrawal at baseline (per TREATDIS.csv), and they don't have diagnosis at baseline.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects in ADNIGO

ADNIGO Study

- 208 subjects from ADNI1(originally enrolled as Normal or MCI) continued to ADNIGO.

  - 208 unique subjects after identified by PHASE='ADNIGO', PTSTATUS=1(Active), RID<2000, and DXCHANGE is not missing in merged REGISTRY/DXSUM/ARM table(N=221 including repeated observations)
  - ADNI1 subjects moved to ADNIGO at month 36, month 48 or month 60.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects in ADNI2

ADNI2 Study (As of July 1 2013)

- New 925 subjects had **screening** observation (Normal:N=262, SMC:N=49, EMCI:N=234, LMCI:N=217, AD:N=163), and N=655 had **baseline** (Normal:N=187, SMC:N=13, EMCI:N=175, LMCI:N=160, AD:N=120).

    - Identified using merged REGISTRY/DXSUM/ARM by PHASE='ADNI2', VISCODE='v01'(screening) or 'v03'(baseline), RGSTATUS=1(Active), and DXCHANGE is not missing.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects in ADNI2

ADNI2 Study (As of July 1 2013)

- ADNI1 subjects: N=258 continued to ADNI2.
- ADNIGO subjects: N=115 continued to ADNI2.

  - Identified using merged REGISTRY/DXSUM/ARM by PHASE='ADNI2', VISCODE='v06'(ADNI2 Initial Visit-continuing Pt), RGSTATUS=1(Active), and DXCHANGE is not missing. (RID=751 has missing on DXCHANGE at v06; telephone visit only)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects: convert from NL to MCI

Create data contains NL to MCI converters using the variables:
DXCHANGE(4:Conv:NL to MCI) and baselineDX=1(NL).

SAS example code to identify NL to MCI converters

```
* sort data first ;
PROC SORT DATA = dxarm_reg;
  BY rid examdate;
RUN;

* output rid and visit time where first time dxchange=4 appeared;
DATA conv_to_MCI(KEEP = rid dxchange phase viscode);
  SET dxarm_reg(WHERE=( dxchange=4 and baselineDX=1));
  BY rid;
  IF FIRST.rid THEN OUTPUT;
RUN;
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Identify subjects: having CSF results

Create data contains subjects who are in the newest data:
UPENNBIOMK6.csv (contains ADNI1,GO,2 results) (Note: VISCODE in
this data is VISCODE2)

SAS example code to merge CSF and dxarm_reg

```
PROC SORT DATA = upennbiomk6;
  BY rid viscode;
PROC SORT DATA = dxarm_reg;
  BY rid viscode2; RUN;

DATA upennbiomk6_dx;
  MERGE dxarm_reg upennbiomk6(RENAME = (viscode = viscode2)) ;
  BY rid viscode2;
  IF abeta=. AND tau=. AND ptau=. THEN DELETE;
RUN;
```

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Cross-validation

What is cross validation?

- Cross validation is a model evaluation method.
- Goal is to avoid over fitting; the model to be generalizable.
- Divide the data into training (to build the model) and test (to evaluate the model) set.
- There are several different ways to validate:
    - Hold-out validation (single train-and-test validation)
    - K-fold cross-validation (a common choice is K=10)
    - Leave-one-out cross-validation (leave one observation out at time; fit the model on the remaining training data)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Cross-validation

CROSSVAL.csv

- Currently, Cross validation file is under Imaging> PET Imaging Analysis.

- The file contains two variables (TRAINING, SET_ID) to separate ADNI1 subjects into partitions.

- For assigning partitions, all ADNI1 subjects were stratified by:
    - Diagnosis: NC, MCI, AD.
    - Study Arm: 1.5T only, PET+ 1.5T, 1.5T and 3T.
    - Young (<76) vs old (>76).

- TRAINING: 40% of subjects are chosen for a training dataset (TRAINING=1), and 60% for a test dataset (TRAINING=0).

- SET_ID: Subjects are divided into 10 parts. (a,b,c,....i,j)

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Cross-validation

Why do we need this file?

- Some imaging labs developed/identified an ROI using subjects in training set, so analysis should focus on subjects in the test set. (TBM.csv and BAINMRC.csv use this approach.)

- Some researchers may want to use these assignments in their own cross validation analysis.

- After enrollment closes for ADNI2, we will be posting similar assignment for ADNIGO2 participants.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Cross-validation

Example: 10-fold Cross-validation

1. Merge ADNI1 data and CROSSVAL.csv using RID.

2. SET_ID variable divides data into 10 parts

3. Set training(9/10) and test(1/10) datasets.

4. Fit the model using training data.

5. Apply the fitted model to the test data.

6. Repeat step 3,4&5 for all 10 sets of the data.

7. Calculate statistics of model accuracy/fit from the test data.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
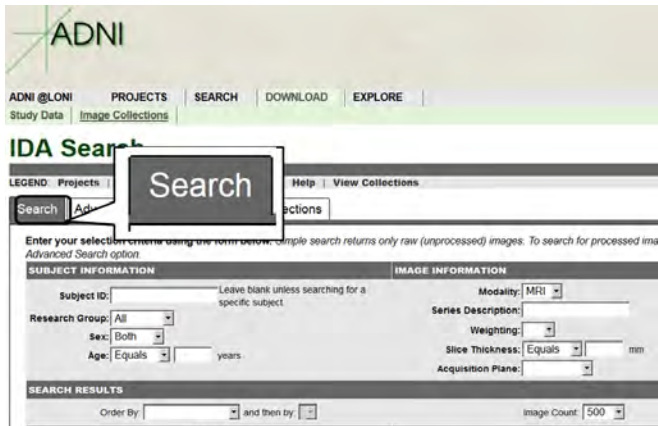
Image Data Archive

The Image Data Archive (IDA) system allows authorized users to download MRI/PET images.
From **Download Study Data** page click **Image Collections**.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Image Data Archive

Search tab: Enables you to search the image database.
This search returns a list of raw images.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Image Data Archive

Search Results tab: Select individual image or **Select All** image.
After you select images, click **ADD TO COLLECTION**.
You can view the images before you download them.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Image Data Archive

Data Collections tab: Choose format and download images.
CSV Download: The file contains image info, age, gender, etc.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Image Data Archive

Under **COLLECTIONS** , you see **Other Shared Collections**
where you can also download collection of images by diagnosis/visit.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data
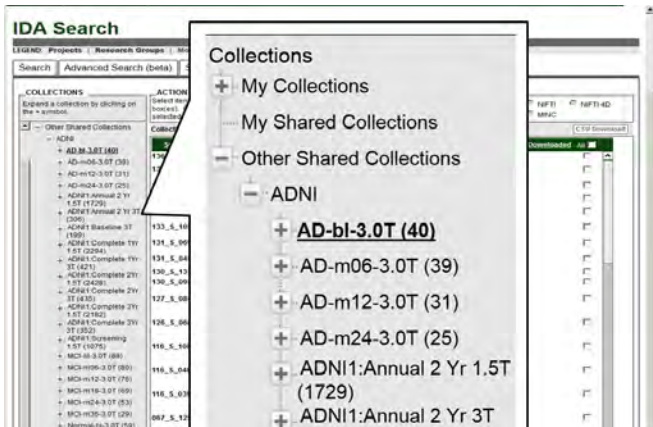
Image Data Archive

Advanced Search(beta): You can search for images by sex, diagnosis, clinical info(MMSE, CDR, NPI, etc), visit(baseline, month6, etc.), and image protocols.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Image Data Archive

IDA manual can be downloaded from **LONI Informatics Core** page.

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Having Question?

If you have questions, please check FAQ section first.

Still no answers?

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Having Question?

Search your question using **Experts Knowledge Base**.

Still no answers?

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Having Question?

Ask the Experts page

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Having Question?

Or you may join ADNI Data User Google Group.
https://groups.google.com/d/forum/adni-data

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Q & A

Any questions?

Data Overview
Commonly Used Data
ADNIMERGE Packages
Working With Data
Cross-validation
IDA Image Data

Thank You

Thank you.