Contents

Data

TADPOLE challenge participants are free to use **any dat** predictions (such as when building predictive models). TADPOLE provides three "standard" data sets, derivec study:

- D1 - a comprehensive longitudinal data set for training;
- D2 - a comprehensive longitudinal data set on rollover subj forecasting;
- D3 - a limited forecasting data set on the same rollover sub

We also refer to D4, the future test set. The archive co standard data and associated files is available from t (login to ADNI, follow Download -> Study Data -> Test

Join

Sign In

Register

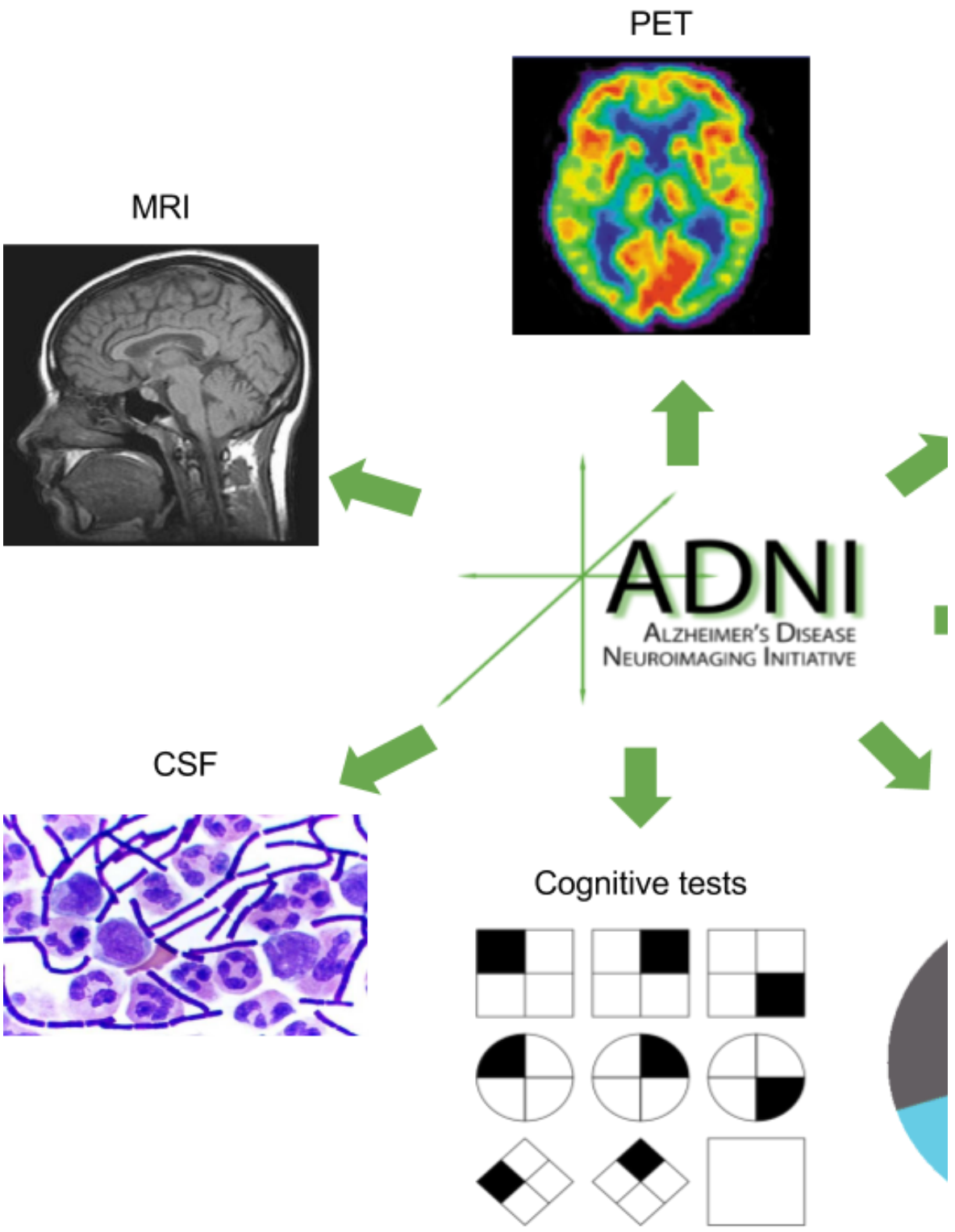


Diagram showing TADPOLE biomarkers. Source of individual i
Commons

Challenges -> "Tadpole Challenge Data"). You will nee
ADNI and apply to access these data files. In the archive,
contains both data sets D1 and D2, with row-wise mem
by a 1 in the "D1" and "D2" columns, respectively. The fi
D3.

All queries on the structure and meaning of the standar
posted on the Google group. The rest of this page prov
content and construction of the TADPOLE standard data s

List of biomarkers

For advancing the diagnosis of dementia, assessmen biomarkers (medical measurements that can indicate a di to cognitive tests is of great value. The five most comm biomarkers were recently included in the revised diagnos and MCI due to AD (Albert et al.,2011; McKhann et al., biomarkers can be divided into two categories: measure beta protein and measures of damage to nerve cells (Jac the first category, amyloid beta can be measur cerebrospinal fluid (CSF) puncture or amyloid p tomography (PET). For the second category, damage to t be measured indirectly by quantifying the fraction of t CSF or using tau-PET, or directly by quantifying brain r fluoro-deoxyglucose (FDG) PET or atrophy using ma imaging (MRI).

TADPOLE standard datasets contain some or all c biomarkers:

1. Main cognitive tests (excluding subtypes): - neuropsycholog administered by a clinical expert

1. CDR Sum of Boxes
2. ADAS11
3. ADAS13
4. MMSE
5. RAVLT
6. Moca
7. Ecog

2. MRI ROIs (Freesurfer) - measures of brain structural integrity

1. volumes

2. cortical thicknesses

3. surface areas

3. FDG PET ROI averages - measure cell metabolism, where cel show reduced metabolism

4. AV45 PET ROI averages - measures amyloid-beta load in the amyloid-beta is a protein that mis-folds (i.e. its 3D structure is constructed), which then leads to AD

5. AV1451 PET ROI averages - measures tau load in the brain, v another protein which, when abnormal, damages neurons and

6. DTI ROI measures - measures microstructural parameters rel axons (cell radial diffusivity, axonal diffusivity, etc ...)

1. Mean diffusivity

2. Axial diffusivity

3. Radial diffusivity

7. CSF biomarkers - amyloid and tau levels in the cerebrospina opposed to the cerebral cortex

8. Others:

1. APOE status - a gene that is a risk factor for developin

2. Demographic information: age, gender, education, etc

3. Diagnosis: either cognitively cormal (CN), mild cogniti (MCI) or Alzheimer's disease (AD).

Getting started

In TADPOLE_D1_D2.csv, each row represents data for one a subject, and each column represents a feature (commonly called biomarker) from the subject at that pa

first columns in the spreadsheet contain unique identifie uniquely identifies every subject, VISCODE (visit code) when the visit takes place (bl is baseline or month 0, m0 ..), SITE represents the site ID where the visit took place columns are: EXAMDATE represents the date of the clir AGE is their age at baseline visit, PTEDUCAT represents t education.

The TADPOLE_D1_D2.csv spreadsheet contains many typ (or measurements), some more important than others start by using only a small subset of the biomarkers whicl informative. Here is a list of biomarkers we suggest parti with ADNI data to start with:

- The main measures to be predicted: DX, ADAS13, Ventricles
- Cognitive tests: CDRSB, ADAS11, MMSE, RAVLT_immediate
- MRI measures: Hippocampus, WholeBrain, Entorhinal, MidT
- PET measures: FDG, AV45
- CSF measures: ABETA_UPENNBIOMK9_04_19_17 (amyloid-l TAU_UPENNBIOMK9_04_19_17 (tau level), PTAU_UPENNBIOMK (phosphorylated tau level)
- Risk factors: APOE4, AGE

Other important biomarkers that participants can consid MRI, PET and DTI measures for the hippocampus, e temporal and parietal lobe structures. Use the (TADPOLE_D1_D2_Dict.csv) and search for keyw "hippocampus" or "hippocampal" to find the necessa example, column ST44CV_UCSFFSL_02_01_16_UCSFFS represents the volume of the left hippocampus. If desir for the left and right structures can be averaged together

Cognitive tests

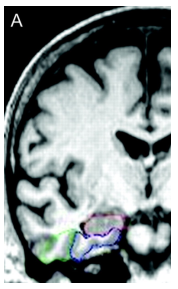
Cognitive tests are neuropsychological tests administered by a clinical expect which assess several skills: general cognition, memory, language, vision, etc ... These cognitive tests give an overall sense of whether a person is aware of their symptoms, is aware of the surrounding environment (i.e. he/she knows where they are, know the date and time) and whether he/she can remember a short list of words, follow instructions and do simple calculations.

Cognitive tests are important in Alzheimer's disease because they measure cognitive decline in a direct and quantifiable manner. In the cascade of pathological events that lead to Alzheimer's disease, cognitive decline is one of the latest to become abnorma the first abnormalities are first noticed on the microscop the misfolding of a protein called Amyloid beta. These changes at larger scales: loss of the neurons' myelin shea visible atrophy in MRI scans and finally cognitive decline. 2010b)

These tests have several limitations: 1. they suffer from p patients who undertake the same test several times car how to do it, and thus score higher at a follow-up vis usefulness of the test in assessing dementia 2. they hav effects, which means that many subjects might score tl score possible and 3. they can be biased, as they are human expert who might be influenced by prior kı subject's cognitive abilities.

MRI measures

Magnetic resonance imaging (MRI) is a technique used to image the anatomy and the physiological processes of the brain and other body parts. With MRI, atrophy can be quantified by measuring the volume of gray matter (GM) and white matter (WM) of the brain. The GM is the brain tissue that consists of nerve cells and the WM consists of fibres connecting these nerve cells. GM can be found in the cortex of the brain and in sub-cortical areas. As a structural MRI scan shows contrast (i.e. differences in pixel intensities) between these tissues, it can be used for volume measurement. Atrophy by indicated by the loss of volume in a particular brain reg scans, one initial scan and one follow-up scan. Atrophy death of neurons in regions affected.



Left: MRI sc onset of atr the subject w AD, which brain. The c deep gray m early ir (hippocar cortex = b green) technology of atrop progressio (AD). Sc

TADPOLE datasets include three main types of structura atrophy: 1. ROI volumes 2. ROI cortical thicknesses 3. F where an ROI (region of interest) is a 3D sub-region of the inferior temporal lobe. Obtaining these structural M the images is a long and complicated process. This involv aligning) the MRI images with each other and performin of the main brain structures using an atlas-based information can be found on the Free https://surfer.nmr.mgh.harvard.edu/fswiki/LongitudinalPr

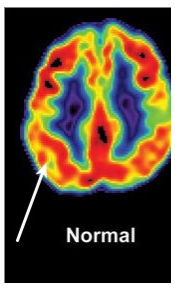
Quantification of atrophy with MRI is a very important widely available and non-invasive. Also, it is a gc progression of MCI to dementia in an individual su becomes abnormal in close temporal proximity to t cognitive impairment (Jack et al., 2013, 2010b).

These measures are computed with an image analysi

Freesurfer using two pipelines: cross-sectional (each independent) or longitudinal (uses information from a subject). The longitudinal measures are ?more robust?, bu that there are more missing values in our TADPOLE spre biomarkers in TADPOLE can be found in the columns co (cross-sectional) and UCSFFSL (longitudinal).

PET measures

Positron Emission Tomography (PET) detects pairs of gamma rays emitted by a radioactive tracer, which is introduced into the body of a biologically active molecule. Three-dimensional images of tracer concentration within the body are then constructed by computer analysis. Before a PET scan, the patient is injected with a contrast agent (containing the tracer) which spreads throughout the brain and binds to abnormal proteins (amyloid and tau). This enables researchers to track the concentration of these proteins. PET scans can be of several types, depending on the cellular and molecular processes that are being measured:



Fluorode images fo subject (le cognitive ir Alzheimer's measures ce is known develop decreased m region Alzheimer's cognitively courtesy o William J

- cell metabolism using Fluorodeoxyglucose (FDG) PET: Neur metabolism refers to the the activity going on inside neuronal processing of food and elimination of waste. Neurons that are show reduced metabolism, so FDG PET is an indicator of neurc PET can be used to measure cell metabolism.

- levels of abnormal proteins such as amyloid-beta through A beta misfolding (i.e. errors in the construction of its 3D structu one of the causes of Alzheimer's disease. High levels of misfolc

in the brain are thought to eventually lead to future neurodege cognitive decline. AV45 PET can be used to measure the levels brain.

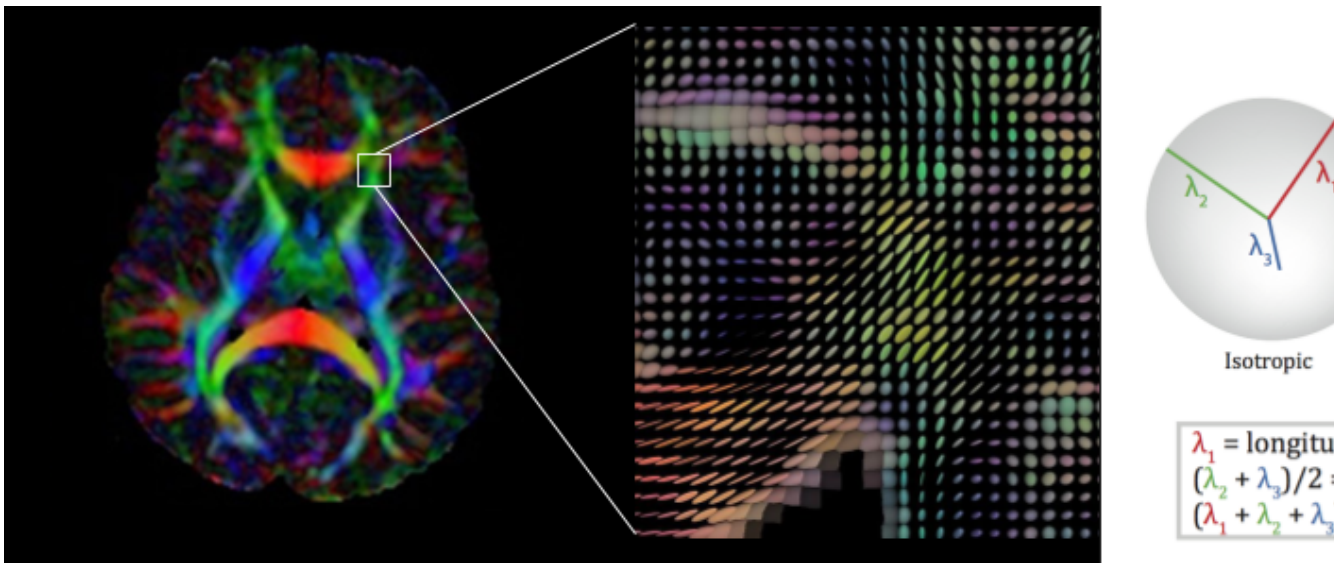
- levels of abnormal tau proteins through AV1451 PET: Abnor phosphorylated tau (i.e. tau protein + a phosphorus group) tha in an insoluble form eventually causes damage to the neuron's causing the neuron's transport system to collapse and thus to

The PET measures are important because they give in molecullar processes that happen in the brain. These are become abnormal in the cascade of events that lea disease, and are therefore important early markers of th about to unfold. In TADPOLE, these PET measures migh whether a healthy control will eventually progress t impairment (MCI) status or not.

While PET scans are non-invasive, they have some limit limitations is that the patient is exposed to ionizing radia the number of scans they can take in a specific time in also have a much lower spatial resolution compared to other caveat with AV1451 PET (tau imaging) is that it is a technology and still under research, and very few subject dataset have undertaken these images. PET measures columns containing "BAIPETNMRC" (FDG PET), "UCBERK and "UCBERKELEYAV1451" (AV1451).

DTI measures

While structural MRI measures brain atrophy, MRI can measure other markers of neurodegeneration that provid information for dementia diagnosis. One such marker is imaging (DTI). DTI can measure the degeneration (connections between neurons) in the brain. This is done diffusion of water molecules along the neuron fi



(Left) Diffusion tensor image of a brain showing white matter fibre co represent the direction of the connection (red for left-right, blue for s green for anterior-posterior). (Middle) Zoomed image into the small re showing the diffusion tensor ellipses. Each ellipse indicates the dire molecules diffused (i.e. moved). (Right) Diagram showing the differen diffusion (i.e. equal in all directions) versus anisotropic diffusion, alon measures that can be computed. Image sources: [1] [2]

Molecular diffusion in tissues is not free, but reflects many obstacles, such as macromolecules, fibers, and me fiber connection degrades, the diffusion becomes more is in every direction), which can be quantified using a fractional anisotropy.

DTI is important for analysing the progression of Alzhe has been shown that dementia affects white matter bu al., 2013). DTI has also shown great potential for aiding dementia (Bozzali et al., 2002; Lu et al., 2014; Zhang et al.,

DTI measures have some limitations. In ADNI, it is a imaging modality, and thus many subjects will not hav Another common problem with diffusion tensor imagi MRI is the partial volume effect, which means that measu (3D pixel) are biased due to averaging across many diffe contained in that voxel. In the TADPOLE spreadsheet, DTI found in columns containing "DTIROI".

CSF measures

The cerebrospinal fluid (CSF) is a clear, colourless body fluid found in the brain and spinal cord. It acts as a cushion or buffer for the brain, providing basic mechanical and immunological protection to the brain inside the skull. A sample of the CSF can be taken from patients invasively, by inserting a needle in the spinal cord, a procedure called lumbar puncture.

Diagra spina blue subara brain

Measures of CSF are very important for dementia researc concentration of abnormal proteins such as amyloid-b strong indicator of AD. Abnormal levels of concentrations are some of the earliest signs of Alzheimer's disease abnormalities many years before symptom onset.

The CSF measures have some limitations. One key limit lumbar puncture is highly invasive and thus not per studies, although a fair amount of ADNI subjects agree procedure. The CSF measures are also not specific to any the brain.

Risk factors

There are several important risk factors that are known to cause dementia. The alipoprotein E4 variant (APOE E4) is a gene that is the largest known risk factor for AD. Subjects with APOE E4 have a risk 10 to 30 times higher of developing AD compared to non-carriers (i.e. subjects without the gene). The exact mechanism through which the presence of APOE E4 leads to AD is not known. The pres

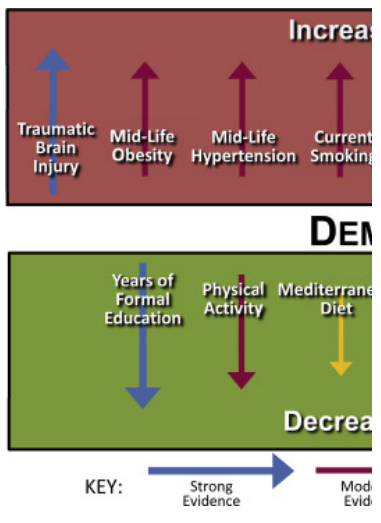


Diagram showing diff to lifestyle and the ass Source: Baum

in a particular subject is denoted by a 1 in the A
TADPOLE_D1_D2.csv

Another known and important risk factor for AD is age – are the more likely they are to develop AD. Above the ag developing dementia doubles every 5 years. Gender is a factor, where women seem more likely to develop AI reasons for this are still unclear.

Finally, there exist many other risk factors related to conditions and lifestyle. Medical conditions such as type blood pressure, high cholesterol, obesity or depressic increase the risk of developing dementia. Lifestyle f increase the risk of developing dementia include p smoking, unhealthy diet, excessive alcohol or head injurie

While some of these risk factors (APOE, age and gender TADPOLE_D1_D2.csv spreadsheet, the other factors are information does however exist in the ADNI database (o under Study Data-> Medical History -> Medical History [TADPOLE participants are welcome to use the inform spreadsheets if desired.

TADPOLE standard data s

The TADPOLE standard data sets can be downloaded fro logging in, go to Download -> Study Data -> Test D Challenges and download "Tadpole Challenge Data".

Here is a description of the TADPOLE standard data sets.

D1. TADPOLE Standard training s

The Standard training set (D1) was created from spreadsheet, to which we added regional MRI (volumes, surface area), PET (FDG, AV45 and AV1451), DTI (re standard indices) and CSF measurements.

The MRI measurements included are FreeSurfer proces cortical thicknesses, and cortical surface areas fro (longitudinal pipeline) and UCSFFSX (cross-sectional Explicitly, spreadsheets UCSF UCSFFSL51ALL_08_01_16.csv, UCSFFSX_11_02_15 UCSFFSX51_08_01_16.csv. Duplicate rows were removed row with the most recent RUNDATE and IMAGEUID.

The PET measurements included are ROI SUVR values AV1451. The spreadsheets used were: BAIPETNM UCBERKELEYAV45_10_17_16.csv, and UCBERKELEYAV1451

The DTI biomarkers included are ROI summary measure spreadsheet DTIROI_04_30_14.csv. For example mean di axial diffusivity AD.

We also included three CSF biomarkers: Amyloid-beta These values were taken from the Elecsys analysis, which the UPENNBIOMK9_04_19_17.csv spreadsheet.

In all cases, we matched rows between ADNIME spreadsheets using the subject ID and visit code. Dup removed, with the most recent preferred. For each r included the ID of the image that was used to derive measures.

D2. TADPOLE Standard predictior

The set of D2 entries contains all currently available lon prospective ADNI-3 subjects that are rollovers from ear Such subjects are active (PTSTATUS=='1'), with (Phase=='ADNI2'), and screening was performed (RGSTA subjects were identified as follows:

1. REGISTRY_ADNI2 = select all from REGISTRY, where Phase== RGSTATUS=='1'

2. DXARM = inner-join of DXSUM and ARM on {RID,Phase} Note: ARM is required for baseline diagnosis. Can also be usec subgroups such as PET+1.5T; 3T+1.5T; etc. (see p23 of ADNI_data_training_slides_part2.pdf)

3. DXARMREG = left-outer-join DXARM and REGISTRY_ADNI2 ({RID,Phase,VISCODE}

4. D2_RID = select RID from DXARMREG, where DXCHANGE is Phase is not missing and PTSTATUS=='1'

5. D2 = historical ADNI data for D2_RID individuals

D3. TADPOLE Cross-sectional predict

D3 uses the same set of participants as D2, but includes and a limited number of data columns. The aim is to min for a clinical trial in which the available information is t demographics, cognitive test scores, and structural M volumes).

D4. TADPOLE Test set

The test set will contain ADNI-3 data from rollover ind after the challenge submission deadline, and used fo forecasts according to the challenge metrics.

Further Details

The TADPOLE standard data sets are downloadable as s the ADNI website. For anyone interested in the det spreadsheets were generated, we have made our scr GitHub. Our repository also contains scripts for leaderboard datasets, and sanity checking and evaluat file.

Organised by: EuroPOND European Progression Of Neurological Disease initiative ADNI

Prize sponsors: alzheimer's association Alzheimer's Society