

## DATA INFORMATION PACK

### EPAD LCS V500.0 data set release

**Author:**

James Howlett

### AUTHORIZATION

Each authorization indicates review and approval of the EPAD LCS Data Information Pack.

**Statistical Team Lead (STL):**

Signature: Signature on file

Date:

Name: Brian Tom

\_\_\_\_\_  
(ddmmmyyyy)

**Sponsor:**

Signature: Signature on file

Date:

Name: Ellie McMaster

\_\_\_\_\_  
(ddmmmyyyy)

**Aridhia Representative:**

Signature: Signature on file

Date:

Name: Harry Peaker

\_\_\_\_\_  
(ddmmmyyyy)

## Data Information Pack (EPAD-LCS)

This document attempts to answer any questions that may arise from researchers analysing the European Prevention of Alzheimer's Disease (EPAD) Longitudinal Cohort Study (LCS). It also documents any decisions taken on what data to include and exclude in each data release.

The data released is as clean and complete as possible at the time of release. Any known issues are documented below. There may be data corrections made subsequently meaning that future releases may contain slightly different information. Any future releases supersede this release. It should be noted that any future changes will be minor.

Aridhia provides a data dictionary which can be accessed at [library.aridhia.net](http://library.aridhia.net).

### 1. Participants included in Vx00.x data releases

The first x00 participants that consent **AND** are entered into the eCRF are included in the Vx00.0 data release. The same participants are included in each follow up data release (Vx00.1, Vx00.2 etc). The situation may arise in Vx00.0 whereby excluded individuals have earlier baseline dates than some included individuals due to the later entry into the eCRF.

### 2. Screen Failures

It is important to note that the data includes every participant consented and that a small proportion of these subsequently failed screening. It may often be appropriate to remove such participants from any analyses performed. Screen failures can be found in the "epadlcs\_discontinuation" table.

### 3. EPAD-LCS Tables

This section gives any information important to know before handling specific EPAD-LCS tables. The names of files containing the data are given in the brackets (prefixed by "epadlcs\_").

#### 3.1. Visit Information (visits)

- The date of visit for each participant at each time point.

#### 3.2. Derived IDs (derids)

- The derived participant ID which links the participant (where applicable) to the parent cohort from which they were invited to join EPAD.

#### 3.3. Consent (consent)

- Informed consent form.

#### 3.4. Eligibility Criteria (eligibility)

- Eligibility criteria met and which criterion not met.

### **3.5. Participant Discontinuation (discontinuation)**

- Participants, dates, and reasons for discontinuation

All discontinuation information available at the time of data release is linked to the last visit that each participant attended. All discontinuation data available for v500 participants with *only one* study visit are provided in v500.0, no matter how long after the visit the participant discontinued. The same logic applies to v500.1 where the participants attended visit 2 or 3.

### **3.6. Demographics (socio\_demographics)**

- Baseline demographic information.

“site\_name” and “site\_id” *does not necessarily correspond* to the baseline site as participants may have their 1<sup>st</sup> visit at one site and then move to another site for subsequent visits. The site given is the site of the participant at the time of the data release. In France, ethnicity data is not allowed to be collected by law, and so for individuals in French sites this information is not available.

### **3.7. Family History of AD (family\_history)**

- Information on each participant’s family history of dementia at baseline.

‘family\_dementia\_history’ is the initial question asked to participants, with family member specific answers recorded only if that family member had a history of dementia. Note that further questions are asked as to whether family members with dementia are related biologically.

### **3.8. Biomarkers (biomarkers)**

- Participant’s APOE genotype from DNA extracted from whole blood with results from each visit for CSF biomarkers.

The Roche CSF assays used have a lower detection limit of 8pg/ml for pTau and 80pg/ml for tTau. Values below the detection limits for pTau and tTau are recorded as <8 and <80 respectively. A $\beta$  has a measuring range of 200-1700pg/ml. Values that are above 1700 are recorded as >1700. These values have been re-calculated to give the actual values and can be extracted from the “abeta\_1\_42\_comments” column. Such values, should be treated with caution as the assay is considered by Roche to be less reliable outside the measuring range.

**In V500.0:** There are 3 participants missing the ‘apoe\_sample\_collected’ variable. APOE samples were not collected for any of these but this fact was not recorded in the database at the time of release.

There are 9 participants with ‘apoe\_sample\_collected’ = ‘Y’. However, there are no APOE results for these participants. This has occurred for a variety of reasons:

- The test was performed but the result was undefined
- Blood samples were taken hence ‘Y’ but no samples for DNA were collected
- Samples are still being followed up

- Samples were collected but lost before being analysed

There were 4 APOE samples which were present in the data but shortly before the release it was made clear that the results from the analysis of the samples were unreliable due to tube handling errors. These samples will be sent for re-testing where appropriate. The current results are not available in v500.0. These samples are all from site\_id 060 (UNIGE). The APOE results for these participants were manually removed by Aridhia. They are recorded as 'apoe\_sample\_collected' = 'Y' but there are no results.

There are 3 participants missing the 'csf\_sample\_collected' variable. CSF samples were not collected for any of these but this fact was not recorded in the database at the time of release.

There are 4 participants where 'csf\_sample\_collected' = 'No' but results are found for these records. CSF was not collected at visit 1 but subsequently was collected at visit 2 (6 months) as a replacement for visit 1. When analysing it should be noted that although the following 'csf\_sample\_id' are linked to visit 1 they were actually collected 6 months later.

- P00355
- P00356
- P00586
- P00587

In one case the CSF sample was linked to the incorrect participant. This has been resolved and *is correct* in the data release. For v500.0 this correction was manually made by Aridhia.

1 participant's baseline CSF was not linked to a visit number. This was manually included by Aridhia.

There are 6 participants with 'csf\_sample\_collected' = 'Y' but have no CSF data. In all cases this has occurred because the CSF samples were collected in the incorrect tubes and therefore were not analysed.

### **ENE Data**

The EPAD Neurological Examination data were included if in the eCRF 'notadmin' = FALSE. Any records where 'notadmin' = TRUE were excluded as the tests were not performed. There was 1 exception (see STAI-40 for details).

The eCRF documents whether each ENE test was performed and the date performed. Any individual without a record in the eCRF is excluded with one exception (see RBANS details).

### **3.9. RBANS (rbans)**

- Primary outcome composite score along with each individual domain and test.

There was 1 participant with no data in the eCRF regarding the RBANS test but valid results were present. This participant was manually included by Aridhia.

Missing data are coded as **995** and should be processed accordingly before any analyses are performed.

### **3.10. TabCat (dot\_counting, favourites, flanker)**

- TabCat tests (dot counting, favourites, and flanker).

These tests were administered using an iPad tablet. It is not currently possible to record in the iPad the visit number the test took place. Only the date the test was administered can be recorded. As such, linking the TabCat data to a visit number is more problematic than for all the other data collected in the EPAD-LCS.

Further information on the EPAD Neurological Examination (ENE) entered into the eCRF allowed the linking of visit number to the TabCat data using the dates provided from the two data sources. Wherever the test data for a participant fell within  $\pm 28$  days of his/her date in the ENE data set, the visit number from the ENE data set was assigned to this participant's test data.

### **3.11. Four Mountains Task (four\_mountains\_uedin)**

- The Four Mountains Task (FMT)

Like TabCat, the test was recorded on an iPad tablet.

However, in addition, the examiner also recorded the answers and the data were entered onto the MedAvante system. Errors were introduced when the examiner recorded the answers and so the data on the MedAvante systems are not provided in the data releases. The MedAvante FMT can be requested separately by researchers wanting to investigate the differences between the two methods of recording FMT.

Linking the test to a visit number was resolved using the same methodology as for TabCat.

There are less data available for the tablet because the iPad storage of the files was unreliable and not all tests were present when the extraction from the iPads was done.

The total FMT score can be calculated by counting the number of 'CORRECT' responses from the 'fms\_uedin\_mark' variables in the data set.

### **3.12. Supermarket Trolley Virtual Reality (vr\_supermarket\_trolley)**

- Supermarket Trolley Virtual Reality (STVR) individual answers and marks.

The total STVR score can be calculated by counting the number of 'Correct' responses from the 'st\_trial\_mark' variables in the data set.

### **3.13. CDR (cdr)**

- Clinical dementia rating global score, sum of boxes and individual domains.

### **3.14. MMSE (mmse)**

- Mini-mental State Examination individual test scores and overall MMSE score.

### **3.15. A-IADL (aiadl)**

- The Amsterdam Instrumental Activities of Daily Living questionnaire.

Missing data are coded as **995** and should be processed accordingly before any analyses are performed.

### **3.16. GDS (gds)**

- Geriatric Depression Scale individual questions and overall score.

Missing data are coded as **995** and should be processed accordingly before any analyses are performed.

### **3.17. PSQI (psqi)**

- Pittsburgh Sleep Quality Index individual items, component scores and overall score.

Missing data are coded as **995** and should be processed accordingly before any analyses are performed.

### **3.18. STAI-40 (stai\_40)**

- State-Trait Anxiety Index individual questions, form scores, and total score.

There was 1 participant for whom the test was not administered according to the eCRF despite a complete set of valid results. This participant was manually added by Aridhia.

Missing data are coded as **995** and should be processed accordingly before any analyses are performed.

### **3.19. Imaging Lacunes and Infarcts (lacunes\_infarcts)**

- Imaging variables related to lacunes and territorial infarcts.

### **3.20. Imaging Radiological Read (radiological\_read)**

- Radiological read imaging variables.

### **3.21. Imaging Volumes (volumetric)**

- Volumetric imaging data.

Many of the volumetric variables are only analysed longitudinally. These variables will be completely missing in the Vx00.0 data releases.

**3.22. Dementia Diagnosis (dementia\_diag)**

- Dementia diagnosis at study visit (only present if dementia was diagnosed by the Investigator), date and dementia type.

**3.23. HATICE Questionnaire (hatice)**

- Healthy Ageing Through Internet Counselling in the Elderly questionnaire.

**3.24. SNAC Questionnaire (snac)**

- Swedish National study on Aging and Care questionnaire.

This data set has multiple records per participant per visit and *is left for the researcher* to process as required.

**3.25. Lifestyle Questionnaire (life)**

- Lifestyle questionnaire on health, activity, smoking and drugs.

**3.26. Physical Examination (physical\_exam)**

- Results from physical examination.

**3.27. Vital Signs (vital\_signs)**

- Vital signs measured at each visit including height, weight, hip and waist circumference, systolic and diastolic blood pressure, and pulse.

**3.28. Adverse Events (adverse\_events)**

- Adverse events information.

Any adverse events that begin before the visit is completed (i.e. within 28 days after visit date) are included. All information available at the time of data release for such adverse events will be included even if the adverse event was resolved after the visit window of 28 days.

**3.29. Current Medication (current\_medication)**

- Current medication information.

Any medication information that is entered in the database before the visit is completed (i.e. within 28 days after visit date) is included. The date of entry into the database is used as current medication is not linked to a particular visit and the date that medication was started is incomplete.

**3.30. Medical History (medical\_history)**

- Medical history information.

Any medical history information is treated in the same way as current medication.

#### 4. DOI (Digital Object Identifier)

Each data set will be registered to a DOI (see table below) for unique and specific identification of the data set in publications and reference materials. This table will be updated with each data release.

Data Set	DOI
V500.0	Doi:10.34688/epadlcs_v500.0_19.05.10