

DATA INFORMATION PACK

EPAD LCS V1500.0 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: 05Dec2019

Name: Brian Tom

(ddmmyyyy)

Sponsor:

Signature: Signature on file

Date: 13Dec2019

Name: Ellie McMaster

(ddmmyyyy)

Aridhia Representative:

Signature: Signature on file

Date: 12Dec2019

Name: Harry Peaker

(ddmmyyyy)

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 Set Library which can be accessed at <https://library.aridhia.net/>. Access to the Dataset Library can be requested by sending an email to EPADLCS@aridhia.com.

Data Release	Data Set Library Tables
V500.0	EPAD LCS DSL V1
V1500.0	EPAD LCS DSL V2

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 participants have earlier baseline dates than some included individuals due to the later entry into the eCRF.

2. Anonymised IDs

The "patient_id" variable included in each table of the V1500.0 data release corresponds to the "patient_id" variable found in V500.0.

3. Changes from V500.0 to V1500.0

Any changes from V500.0 to V1500.0 that affect individual tables are included in the relevant sections below.

3.1. Table Changes

The "epadlcs_four_mountains_medavante" table is now included. The "epadlcs_biomarkers" table is longer included in the data release. This has been replaced by the "epadlcs_apoe" and "epadlcs_csf" tables.

In future releases an additional table "epadlcs_mri_scanner_information" will be included containing details of the scanner in which the MRI was performed. It was decided not to release this for V1500.0 as this information only becomes relevant to researchers if a participant changes scanner from between visits. This information will therefore be released in longitudinal data releases.

3.2. Additional Variables

With the exception of the “epadlcs_derids”, “epadlcs_consent”, “epadlcs_socio_demographics”, “epadlcs_family_history”, “epadlcs_discontinuation”, “epadlcs_adverse_events”, “epadlcs_current_medication” and “epadlcs_medical_history” tables a “visit” variable has been added to all tables. In V1500.0 this will always be V1 but this variable will be important for future data releases containing longitudinal data.

Table	Variable Name
visits	visit_site
favourites	fav_learn_r1_int
favourites	fav_learn_r2_int
favourites	fav_delay_int
favourites	fav_delay_total_int
favourites	fav_rec_dprime
favourites	fav_rec_bias

3.3. Variable Name Changes

Table	Old Variable Name	New Variable Name
rbans	reason_not_done	reason_not_performed
dot_counting	trial_*_correct	dotcount_t*
dot_counting	total_score	dotcount_total
dot_counting	good_for_research	valid_data
dot_counting	admin_comments	examiner_comments
favourites	recall_one_*	fav_learn_r1_*
favourites	recall_two_*	fav_learn_r2_*
favourites	delayed_recall_*	fav_delay_*
favourites	rec_trial_*	fav_rec_t*
favourites	rec_total_correct	fav_rec_total_correct
favourites	rec_sme	fav_rec_total_sme
favourites	rec_intrusions	fav_rec_total_int
favourites	good_for_research	valid_data
favourites	admin_comments	examiner_comments
flanker	total_*	flanker_total_*
flanker	congr_*	flanker_congr_*
flanker	incongr_*	flanker_incongr_*
flanker	left_*	flanker_left_*
flanker	right_*	flanker_right_*
flanker	up_*	flanker_up_*
flanker	down_*	flanker_down_*
flanker	good_for_research	valid_data
flanker	admin_comments	examiner_comments
vr_supermarket_trolley	reason_not_done	reason_not_performed
cdr	reason_not_done	reason_not_performed

mmse	mmse_Calculation_*	mmse_calculation_*
mmse	mmse_Comprehension_*	mmse_comprehension_*
mmse	reason_not_done	reason_not_performed
aiadl	a17b	q17b
aiadl	a21b	q21b
aiadl	a38a	q38b
psqi	psqi_component*_score	psqi_component*_score
stai_40	reason_not_done	reason_not_performed
dementia_diag	body_system_organ_or_class	body_system_or_organ_class
snac	lcs_3_raw	lcs23_3_raw
medical_history	lowest_level_code	lowest_level_term_code

3.4. Individual Data Changes

Between the V500.0 and V1500.0 releases, some data have been corrected resulting in differences between the two releases.

- “age_years” and “age_months” has been corrected in the “socio_demographics” table for 1 participant
- “handedness” has been corrected in the “socio_demographics” table for 1 participant
- 1 site had previously misunderstood the “bio_relative” questions in the “family_history” table. As a result many responses have been corrected from “No” to “Yes” for this site.
- “mother_age_at_diagnosis”, “father_age_at_diagnosis”, and “brother_age_at_diagnosis” have been corrected in the “family_history” table for 3, 1, and 2 participants respectively
- 2 participants did not have “apoe_sample_collected” and “apoe_reason_not_collected” available in V500.0 but are now included in V1500.0 “apoe” table
- “apoe_sample_date” has been updated for 2 participants in the “apoe” table
- 2 participants did not have “csf_sample_collected” and “csf_reason_not_collected” available in V500.0 but are now included in V1500.0 “csf” table
- “csf_reason_not_collected” and “csf_sample_date” have been updated for 1 and 4 participants respectively in the “csf” table
- “csf_sample_id” has been updated due to data reconciliation processes
- “csf_sample_collected” was “N” in V500.0 but are now “Y” in V1500.0 for 4 participants. These 4 participants are now included as historical samples were available and the EPAD Lite process was followed. “csf_reason_not_collected” has also been updated for these participants in the “csf” table
- 2 participants did not have any information available in V500.0 on whether the assessment was performed, the assessment date and any reasons for not being collected for “rbans”, “dot_counting”, “favourites”, “flanker”, “vr_supermarket_trolley”, “cdr”, “mmse”, “aiadl”, “gds”, “psqi”, and “stai_40”. These are included in V1500.0
- 3 participants had completely missing “dot_counting” records in V500.0 but are present in V1500.0 as a result of data cleaning processes

- “link_date” and “assessment_date” have been updated for 2 and 1 participants respectively in the “dot_counting” table
- 1 participant had completely missing “favourites” records in V500.0 but are present in V1500.0 as a result of data cleaning processes
- “assessment_performed”, “assessment_date”, and “reason_not_performed” have been updated for 1 participants in the “favourites” table
- “link_date” and “assessment_date” have been updated for 2 and 1 participants respectively in the “favourites” table
- “fav_delay_food_1_score” has been updated for all participants in the “favourites” table
- 3 participants had completely missing “flanker” records in V500.0 but are present in V1500.0 as a result of data cleaning processes
- “link_date” and “assessment_date” have been updated for 2 and 1 participants respectively in the “dot_counting” table
- “assessment_performed”, “assessment_date”, and “reason_not_performed” has been updated for 1 participant in the “flanker” table
- 35 participants had completely missing “four_mountains_uedin” records in V500.0 but are present in V1500.0 as a result of data cleaning processes. In addition, 1 participant had records in V500.0 but are now missing in V1500.0
- 1 participant did not have “dementia_diagnosed” available in V500.0 but is now included in V1500.0 “dementia_diag” table
- In the “dementia_diag” table, the values are now character strings rather than numeric codes.
- 1 participant did not have “lifestyle_assessment_completed” and “reason_not_completed” available in V500.0 but is now included in V1500.0 “life” table
- 1 participant did not have “assessment_performed” and “reason_not_done” available in V500.0 but is now included in V1500.0 “hatice” table
- 1 participant did not have “lcs12_5” available in V500.0 but is now included in the V1500.0 “hatice” table
- 1 participant did not have “vital_signs_collected” and “reason_not_collected” available in V500.0 but is now included in V1500.0 “vital_signs” table
- “pulse” has been updated for 1 participant in the “vital_signs” table
- 1 participant did not have “was_physical_exam_performed” and “reason_not_performed” available in V500.0 but is now included in V1500.0 “physical_exam” table

3.5. Imaging Data Changes

IXICO provided the following statement as to why the imaging results may be slightly different in V1500.0 to V500.0.

Radiology Reads

The central Radiology Read process allows the reader to review the previous imaging visits' scans during the following read, to facilitate completion of fields which request 'changes from previous visit' to be reported.

During this review there are instances where the reader notices minor findings that were missed in the initial review. In this case the reader requests the previously submitted read report and updates the relevant fields.

Most typically this will be a re-count of microbleeds which does not change the clinical significance of the report.

When the updated report is re-submitted by the readers an updated read notification is sent to sites and the data transfers to Aridhia will be updated accordingly.

If subject doesn't have an eligibility read (for example if some mandatory sequences are missing or of very poor quality) that subject will not have any follow up reads (as the follow up read report is dependent on comparison against the eligibility scan).

Volumetric Results

The volumetric analysis in EPAD is run when the scans arrive at IXICO. The analysis is automated but is completed with a visual endpoint quality check (QC) done by trained image analysts.

Prior to the monthly data transfers the volumetric data also undergoes a science review in which the quantitative data are checked for completeness and outliers. Where there are outliers the Lead EPAD Biomarker Scientist will ask the Image Analysis team to re-review the analysis. In some instances, the endpoint QC result may change based on re-review from a pass to a fail or vice versa.

4. 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.

5. 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_").

Missing data are detailed in the individual sections relating to each table. In addition, 1 participant included in V1500.0 has no data entered into the eCRF. This participant was a screen failure and shortly after completing consent the site closed. It is uncertain whether this participant's data will be available in future data releases. Any further missing data comments below, are in addition to this participant.

5.1. Visit Information (visits)

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

5.2. 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.

5.3. Consent (consent)

- Informed consent form.

40 participants have no data on “date_study_partner”. Most often this occurs as the participant fails screening before the study partner consent was signed.

5.4. Eligibility Criteria (eligibility)

- Eligibility criteria met and which criterion not met.

2 participants do not have eligibility data. This has been queried but these data will not be available at the time of the V1500.0 release.

5.5. 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.

5.6. Demographics (socio_demographics)

- Baseline demographic information.

“site_name” and “site_id” *does not necessarily correspond* to the baseline site as participants may have their 1st 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.

There are 31 participants with missing data in the “socio_demographics” table. 21 participants are only missing “handedness” data. Of these, 2 use both hands equally and 19 are unknown and are now withdrawn from the study and so will never be known.

1 participant has no demographic data. This has been queried but will not be available at the time of the V1500.0 data release.

9 participants have a combination of “handedness”, “years_education”, “marital_status” and “ethnicity” missing. These participants have all discontinued and will not ever be available.

The eCRF incorrectly recorded “age_years” for 1 participant. Aridhia have manually corrected this age for this data release.

5.7. 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. The response to individual family member dementia history questions are only recorded if the answers to “family_dementia_history” and the individual family member dementia history are “Yes”. If the answer to “family_dementia_history” is “No” or if the answer to “family_dementia_history” is “Yes” *and* the individual family dementia history question is missing it can be assumed that the individual family member does not have a history of dementia. If an individual family member has a history of dementia further questions are asked to determine whether the family member was a biological relative and the age of dementia diagnosis. It is possible for the same type of family member to have multiple dementia history (e.g. two sisters both having a history of dementia). These responses are recorded in “sister2_dementia_history”, “sister2_bio_relative” etc.

10 participants are missing “family_dementia_history”.

5.8. APOE (apoe)

- Participant’s APOE genotype from DNA extracted from whole blood

3 participants have no information as to whether the APOE sample was collected. 2 of these did not have blood samples taken. 1 participant used visit 1 kits at visit 2. An APOE result can be expected for this participant in future data releases.

125 participants are recorded as having an APOE sample collected for whom there are no results in the database. 69 of these samples have been returned by the site and have been sent for testing with the results expected in future data releases. 16 sample results have already been released and will be available in future data releases. 7 samples have recently been returned by the site and will be available in future data releases. 22 samples have been recorded as having been taken but have not yet been returned by the site. Some of these samples may never be returned but some results will be available in future data releases. 3 samples were taken but were destroyed or discarded at the site. 5 blood samples were taken but DNA samples were not collected. 3 of these results may be available in future releases if the DNA can be collected at a future visit. 3 samples have been incorrectly labelled as a visit 3 and so have yet to be sent for testing. These results may be available in future data releases.

5.9. CSF (csf)

- Participant’s results from each visit for CSF biomarkers.

The Roche CSF assays used have a lower detection limit of 200pg/ml for A β , 8pg/ml for pTau and 80pg/ml for tTau. Values below the detection limits for A β , pTau and tTau are recorded as <200, <8 and <80 respectively. A β 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.

Roche give the following disclaimer about A β values >1700:

The Elecsys β -Amyloid (1-42) CSF immunoassay in use is not a commercially available IVD assay. It is an assay that is currently under development and for investigational use only. The measuring range of the assay is 200 (lower technical limit) – 1700 pg/mL (upper technical limit). The performance of the assay beyond the upper technical limit has not been formally established. Therefore, use of values above the upper technical limit, which are provided based on an extrapolation of the calibration curve, is restricted to exploratory research purposes and is excluded for clinical decision making or for the derivation of medical decision points.

14 participants did not have CSF taken at Visit 1. Instead, they were taken at Visit 2 as a baseline measurement. These individuals can be identified using the “csf_retest”, “csf_retest_reason”, and “csf_retest_visit” variables.

5 participants have no information as to whether the CSF sample was collected. For three of these participants, CSF was not collected; for 1 participant it is unknown; and 1 participant had CSF collected and results should be available in a future data release.

There are 46 CSF samples recorded as having been taken where CSF results are not given and no “reason_not_analysed” provided. 12 samples have been returned and most results should be available in future data releases. 19 samples are still to be returned and results should be available in future data releases. 12 samples have become available through the **EPAD-Lite** process and will be available in future data releases. 1 sample was collected but discarded by the site. 1 kit was returned without the CSF samples. 1 participant’s CSF results should become available as a result of data cleaning processes. 1 “csf_sample_date” is missing.

Multiple CSF results were associated for 1 CSF sample ID. Aridhia have manually removed the incorrect results for this data release.

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.

In addition to the participant who has no data entered into the eCRF, 1 individual has no data entered for any of the ENE data.

5.10. RBANS (rbans)

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

Missing data are coded as **995** and should be processed accordingly before any analyses are performed. 10 individuals did not complete individual RBANS tests and therefore, not all of the indices and totals can be calculated. 1 individual did not complete any of the RBANS tests despite the test being opened.

5.11. 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 ± 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.

The linkage problem between the eCRF and TabCat data is further compounded by the fact that wrongly entered participant ID into either TabCat or eCRF could result in a mismatched record. The data cleaning process resolved most of the cases that were identified as having record mismatch. However, there were issues with 83 participants that were unable to link the two sources of data. Further investigation is being carried out and therefore these participants' TabCat data will not be part of the current data release. We hope to make these data available in future update release. Furthermore, 141 participants have missing TabCat data due to synchronisation issues between the data collection iPad and the server where the data is stored.

The eCRF incorrectly recorded that 1 participant did not have the TabCat battery of tests for visit 1. Aridhia has manually corrected this for V1500.0. In addition the eCRF incorrectly recorded the date on which the favourites was performed for 2 participants. Aridhia has manually corrected this for V1500.0.

5.12. 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.

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. Efforts are being made to retrieve the rest the data from the iPads. Any additional data that will be retrieved will be made available as part of future update release. There are 712 participants where the eCRF records FMT as being done but there are currently no iPad records available.

The total FMT score can be calculated by counting the number of “CORRECT” responses from the “fms_uedin_mark” variables in the data set.

The eCRF incorrectly recorded that 1 participant did not have the FMT for visit 1. Aridhia has manually corrected this for V1500.0. In addition the eCRF incorrectly recorded the date on which the FMT was performed for 2 participants. Aridhia has manually corrected this for V1500.0.

5.13. Four Mountains Task (four_mountains_medavante)

- The Four Mountains Task (FMT)

It was previously believed that Four Mountains Medavante data were unreliable compared to the iPad recorded test. However, after the data release it was discovered that the issue was with data processing and this has now been resolved.

There are 43 individuals where the FMT test was done but some of the responses are missing.

The manual changes made by Aridhia for “four_mountains_uedin” also apply here.

5.14. 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.

Where the STVR was performed according to the eCRF there are 21 records with missing data in some of the fields.

5.15. CDR (cdr)

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

There is 1 record where CDR is partially missing. There are 7 individual where CDR was performed but have completely missing records.

5.16. MMSE (mmse)

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

5.17. 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.

5.18. 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.

5.19. 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 PSQI records are completely missing despite the assessment being performed according to the eCRF.

Manual recoding of 1 participant's answers to the PSQI items was performed by Aridhia to ensure consistency of coding across all participants.

5.20. STAI-40 (stai_40)

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

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

There are 2 STAI-40 records completely missing where the assessment was performed according to the eCRF. In addition, 11 records are partially missing.

5.21. Imaging Lacunes and Infarcts (lacunes_infarcts)

- Imaging variables related to lacunes and territorial infarcts.

100 participants did not have an MRI scan and so have no data. In addition, results for 2 participants were re-opened and so are not available in this data release.

5.22. Imaging Radiological Read (radiological_read)

- Radiological read imaging variables.

100 participants did not have an MRI scan and so have no data. In addition, results for 2 participants were re-opened and so are not available in this data release.

5.23. 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.

100 participants did not have an MRI scan and so have no data. A further 10 images were unable to be processed to obtain volumetric results.

5.24. Dementia Diagnosis (dementia_diag)

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

6 participants have no information as to their dementia diagnosis. In all instances these participants failed screening. 1 participant has no "date_of_diagnosis" collected.

5.25. HATICE Questionnaire (hatice)

- Healthy Ageing Through Internet Counselling in the Elderly questionnaire.

6 participants have no information as to whether the HATICE assessment was performed. 5 of these participants either failed screening or withdrew consent. The other participant's data may be available in future data releases. A further 19 participants are partially missing HATICE as some of the questions were not answered. These data will never be available.

5.26. 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.

5.27. Lifestyle Questionnaire (life)

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

3 participants have no information as to whether the lifestyle questionnaire was completed. All 3 are no longer enrolled in the study and these data will never be available. A further 2 participants are partially missing lifestyle information.

5.28. Physical Examination (physical_exam)

- Results from physical examination.

23 participants have no information in "was_ecg_performed". These are generally screen failures and the results will never be available.

5.29. Vital Signs (vital_signs)

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

4 participants have no vital signs collected and are all screen failures meaning no data will be available for these participants. A further 26 participants have partially missing vital signs.

5.30. 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.

5.31. 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.

5.32. Medical History (medical_history)

- Medical history information.

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

6. Other Information

When analysing the EPAD-LCS data consideration should be given to the sampling method used. Details on the participant selection process by the EPAD-LCS Balancing Committee can be found in the EPAD-LCS protocol.

The Brain Injury Screening Questionnaire (BISQ) is not included in this data release as the data are currently unavailable.

Researchers can apply for access to the biological samples or MRI Image data via the ERAP website: www.ep-ad.org/erap.

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
V1500.0	Doi:10.34688/epadlcs_v1500.0_19.11.29