

DATA INFORMATION PACK

EPAD LCS V.IMI data set release of 30th October 2020

Author:

James Howlett

AUTHORIZATION

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

Statistical Team Lead (STL):

Signature: On file

Date: 30Oct2020

Name: Brian Tom

(ddmmyyyy)

Sponsor:

Signature: On file

Date: 30Oct2020

Name: Neil Mitchell

(ddmmyyyy)

Aridhia Representative:

Signature: On file

Date: 30Oct2020

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. All issues arising have been dealt with appropriately, whether they have been resolved pre- or post- locking of the IQVIA's EPAD database on 26th August 2020 or documented as have been queried but unresolved due to source data at Trial Delivery Centres being unavailable because of COVID-19. 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 corrections to the V.IMI data set should be minor. Additional data generated after the IMI period of EPAD will be associated and linked to the V.IMI data set.

Aridhia provides FAIR Data Services which can be accessed at <https://fair.addi.ad-datinitiative.org/#/data/home>

All abbreviations used in this DIP can be accessed in the FAIR Data Services.

Data Release	FAIR Dataset
V500.0	EPAD LCS v500.0
V1500.0	EPAD LCS v1500.0
V500.1	EPAD LCS v500.1
V.IMI	EPAD LCS v.imi

1. Participants included in earlier 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. Participants included in V.IMI data release

The 2096 participants that consented and were entered into the eCRF during the IMI-period of funding are included in the V.IMI data release of the 30th October 2020.

3. Anonymised IDs

The "patient_id" variable included in each table of the V.IMI data release corresponds to the "patient_id" variable found in previous data releases.

4. Changes from V1500.0/V500.1 to V.IMI

Any changes from V1500.0/V500.1 to the 30th October 2020 data release V.IMI that affect individual tables are included in the relevant sections below.

4.1. Variable Name Changes

Table	Old Variable Name	New Variable Name
visits	reason_not_done	reason_not_performed
aiadl	reason_not_done	reason_not_performed
gds	reason_not_done	reason_not_performed
psqi	reason_not_done	reason_not_performed
hatice	reason_not_done	reason_not_performed

4.2. Individual Data Changes

Between the V1500.0/V500.1 and V.IMI releases, some data have been corrected resulting in differences between these releases

4.2.1. Individual Data Changes since V500.1

- 3 participants' records in the "visits" table have been removed as the visit did not take place
- Eligibility criterion not met has been updated for 1 participant in the "eligibility" table
- Years education has been updated for 1 participant in the "socio_demographics" table
- APOE sample information has been updated for 1 participant in "apoe"
- CSF sample information has been updated for 1 participant in "csf"
- 2 records with previously missing CSF results are now included in "csf"
- 40 new records have been included in the "dot_counting" table
- 40 new records have been included in the "favourites" table
- 2 records previously available are no longer included
- "fav_learn_r1_int", "fav_learn_r2_int", "fav_delay_int", and "fav_delay_total_int" have been updated for 120 records
- "assessment_performed", "assessment_date" and "reason_not_performed" have been updated for 4 "dot_counting", "favourites" and "flanker" records
- 119 new records have been included in the "four_mountains_edinburgh" table
- 'mark' variables in "four_mountains_edinburgh" can now take additional values of 'DELAYED_CORRECT', 'DELAYED_INCORRECT' and 'NO_RESPONSE'. 'answer' variables can now additionally take values of 'NO_RESPONSE'
- Assessment date has been updated for 1 participant in the "four_mountains_edinburgh", "four_mountains_medavante" and "four_mountains_tabcat" tables
- "assessment_performed" and "reason_not_performed" has been changed for 3 records in the "vr_supermarket_trolley_medavante" and "vr_supermarket_trolley_tabcat" tables. "assessment_date" has been updated for 4 records
- "cdr_global_score" and "cdr_sum_of_box" have been updated for 1 participant where they were previously missing in the "cdr" table
- 2 participants' values have been entered where they were previously missing in the "psqi" table
- 1 participants' values have been entered where they were previously missing in the "stai_40" table

- 6 participants' imaging measures across 8 visits were previously missing have now been included
- 1 participant previously had imaging results but these are no longer included
- A number of imaging measures have been updated – see statement from IXICO below
- The “dementia_diag” table has been updated for 1 participant
- Discontinuation date has been updated for 1 participant and discontinuation reason has been updated for 2 participants in the “discontinuation” table
- Date of vital signs has been updated for 1 participant in the “vital_signs” table

4.2.2. Individual Data Changes since V1500.0

- 1 record has been included in the “visits” table that was previously missing
- “visdat_int” has been updated for 2 records in the “visits” table
- The “eligibility” table has been updated for 20 records
- 1 record has been included in the “socio_demographics” table that was previously missing
- “age_months” is no longer available for Swiss sites in the “socio_demographics” table. It has also been updated for a further 4 records. Changes to the date of birth has also resulted in minor changes to the calculation of “age_years”.
- “handedness” has been updated for 1 participant in “socio_demographics”
- “years_education” has been updated for 3 participants in “socio_demographics”
- “sex” has been updated for 1 participant in “socio_demographics”
- “apoe_sample_date” has been updated for 1 participant and “apoe_blood_sample_id” has been updated for 1 participant in “apoe”
- 89 “apoe_result” records that were previously missing are now included in “apoe”
- CSF sample information has been updated for 8 participants in “csf”
- 28 CSF results that were previously missing are now included in “csf”
- “rbans” assessment information has been updated for 2 participants
- “dot_counting” assessment information has been updated for 14 participants. In addition, 139 records previously missing have been included
- “favourites” assessment information has been updated for 13 participants. In addition, 141 records previously missing have been included
- “flanker” assessment information has been updated for 14 participants
- “answer*” variables are now coded T, C, S, E rather than TL, TR, BL, BR in “four_mountains_edinburgh”
- 2 “assessment_date” records have been updated in “four_mountains_edinburgh”
- 308 records previously missing have been included in “four_mountains_edinburgh”
- “answer*” variables are now coded T, C, S, E rather than 0, 1, 2, 3 in “four_mountains_medavante”
- Assessment information has been updated for 3 participants in “four_mountains_medavante”

- Assessment information has been updated for 12 participants in “vr_supermarket_trolley_medavante”
- Assessment information has been updated for 1 participant in “cdr”, “mmse”, “aiadl” and “gds”
- Assessment information has been updated for 3 participants in “psqi”
- Assessment information has been updated for 4 participants in “stai-40”
- 918 records that were previously missing have been included in “mri_scanner_information”
- 4 records that were previously missing have been included in “lacunes_infarcts”
- “AHMICLC” was previously missing but is now included in “radiological_read”
- 4 records that were previously missing have been included in “radiological_read”
- 5 records that were previously missing have been included in “volumetric”
- “dementia_diag” has been updated for 1 participant
- “discontinuation” has been updated for 6 participants
- “life” has been updated for 1 participant
- “date_collected” has been updated for 1 record in “vital_signs”
- 2 records that were previously missing have been updated in “hatice”

4.3. Imaging Data Changes

IXICO provided the following statement as to why the imaging results may be slightly different between data releases.

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.

5. 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. Careful examination of this table will allow an informed decision to be made with regard to the participants that should be removed.

6. 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.

6.1. Visit Information (visits)

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

6.2. 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.

6.3. Consent (consent)

- Informed consent form.

6.4. Eligibility Criteria (eligibility)

- Eligibility criteria met and which criterion not met.

25 participants have no information as to which eligibility criterion was not met

6.5. Participant Discontinuation (discontinuation)

- Participants, dates, and reasons for discontinuation

6.6. 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 between subsequent visits. The site given is the site of the participant at the time of the data release. In France, ethnicity data is not permitted to be collected by law, and so for individuals in French sites this information is not available. In Switzerland, only the year of birth is permitted to be collected by law, and so for individuals in Swiss sites no information is available on “age_months”.

There are 33 participants with missing data in the “socio_demographics” table. 20 participants are only missing “handedness” data. This is either because the information is unknown and the participant has withdrawn from the study or the participant uses both hands equally. 2 participants are only missing “years_education”. 1 participant is only missing “marital_status”.

2 participants are missing “handedness” and “years_education”. 1 participant is missing “handedness” and “ethnicity”. 3 participants are missing “years_education” and “marital_status”.

1 participant is missing “handedness”, “years_education” and “ethnicity”. 2 participants are missing “handedness”, “years_education” and “marital_status”.

1 participant is missing all demographic information except “site_name” and “site_id”.

6.7. 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* there is no record of a particular family member it can be assumed that the particular 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).

14 “family_history” records have no information in “family_dementia_history” and 188 “family_history” records are missing “age_at_diagnosis”.

6.8. APOE (apoe)

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

6 participants have no information as to whether an APOE sample was collected. In addition, 58 participants had an APOE sample collected according to the eCRF but have no APOE result available.

6.9. 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.

144 records have no information as to whether a CSF sample was collected. In addition, 27 participants had a CSF sample collected according to the eCRF but have no CSF results available.

18 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.

The EPAD Balancing Committee used the following definition for amyloid status when monitoring the cohort characteristics over time and dynamically deciding on the recruitment strategy to employ.

- Amyloid positive: A β < 1000pg/ml
- Amyloid negative: A β > 1200pg/ml
- Amyloid status undecided: 1000pg/ml \leq A β \leq 1200pg/ml

6.10. 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.

EPAD Neuropsychological Evaluation (ENE) implementation, management and data collection in the EPAD LCS

TabCAT is an iOS application for neuropsychological testing developed and managed by University California San Francisco, Memory and Ageing Centre with an on-line portal for trial set-up, administration and data management. Three tests from the TabCAT battery were selected for the

EPAD ENE: Dot Counting, Flanker and Favourites. The tests are administered using an iPad supported by an Examiner who guides the participant through the tests.

- Dot Counting is an Examiner recorded test,
- Flanker is a Participant recorded test,
- Favourites has three parts; Learning and Delay are Examiner recorded tests and Recognition is a participant recorded test.

This means that source documents are created at the Trial Delivery Centre (TDC) for the Examiner recorded tests which the Examiner uses to enter the data into the TabCAT app. Participant recorded tests can only be recorded directly by the participant in the TabCAT app and the only record of participant recorded data is held in TabCAT. Instruction screens were available in English and Spanish when the study was opened in May 2016. Over time validated translations for other required languages were added to the system. Tests were not implemented without locally appropriate translations, as approved by the relevant ethics committee, therefore some data missing are from some TDCs early in the study and the tests are recorded as Not Done in the eCRF.

There have been several technical and logistical challenges during the EPAD LCS which have impacted on data collection with TabCAT app. The process for synchronising the data recorded in the TabCAT app with the data server (at UCSF) has at times been problematic. A stable internet connection is required for the sync process to be successful and some hospitals/sites have logistical issues getting access to good internet connection suitable for use with an iPad supplied separately from hospital/site equipment. Also, TDC staff have not always understood the sync requirements or been diligent in ensuring this is done at least once per day before the app/iPad is closed down. These factors have resulted in participant visits (Encounters) being lost. These Encounters have been identified, the eCRF records which tests were done, and the TDCs have, where possible, located the source documents from Examiner recorded tests, and have re-entered the data into TabCAT. The eCRF records have not been changed - usually all three tests will be reported as Done, but if the data were lost in the sync process and then re-entered from source documents, there will only be a record of data for Dot Counting and Favourites Learning and Delay. All valid data that can be correctly matched to an eCRF entry have been included - if a test is marked as "Done" in the eCRF but is not present in the data set then the reason for this will be one of those explained above.

The TabCAT app in use at the start of the study included a Demographics Screen, which Examiners were instructed not to use as these data were collected in the eCRF. It was identified in late 2018 that some demographics data had been entered by some Examiners at some TDCs so personal information was recorded in the data base available to Data Management staff and TabCAT administrators. The CI and Sponsor requested that the system was switched off until this issue was corrected. No data were collected for the 3 TabCAT tests, Dot Counting, Flanker and Favourites between 20th Dec 2018 and 12th Sep 2019.

Four Mountains (FMT) test and Supermarket trolley (SMT) tests are also included in the EPAD ENE. Collection of these two tests was not impacted by the TabCAT halt. The introduction of the tests was delayed in some countries pending provision of approved translations of instruction

screens. These two tests were provided as iOS applications from separate sources/licenses not associated with the TabCAT app. From the beginning of the study until 12th Sep 2019 these tests were administered by Examiners who observed and recorded the answer responses of the participants on a paper score sheet. The answer responses were then entered into container forms (data entry screens) on the Virgil Tablet provided to EPAD by MedAvante. In addition to this Examiner collected data, the FMT app creates a copy of the answers (as tapped on the screen by the participant and observed by the Examiner) and the reaction time for each answer directly into a .txt file on the iPad. It is important to collect reaction time in these type of tests as an indicator of cognitive decline when measured over time. There were also technical and logistical issues with data collection of FMT using the iPad app which resulted in missing data in the .txt file when the eCRF record is completed as "Done" and there is a record in the MedAvante Examiner collected data. If the Participant Number field in the FMT app was not completed by the Examiner administering the test then a .txt record was not created; this has resulted in some missing data. The .txt file can only be downloaded by connection with the iTunes app. This process was deemed to present too many logistical challenges to be rolled out to TDCs, therefore the FMT iPad data was collected centrally by periodically returning the iPads to UEDIN staff for data extraction. MedAvante took on a logistics management role for the iPad provisioning and updates which involved using security software to prevent TDCs being able to make changes to the iPad settings. Some data were also lost in the extraction process from these iPads due to faulty signalling from the security software which resulted in the iPads being reset to factory conditions which also wiped the FMT .txt files. All valid data that can be correctly matched to an eCRF entry have been included - if a test is marked as "Done" in the eCRF but is not present in the data set then the reason for this will be one of those explained above.

TabCAT was upgraded during the study to introduce a clinical trial management system that allowed corrections and audit trail collection for all actions in TabCAT and the demographics screen was modified to be an option on a per study basis. In addition, the FMT and SMT were configured as new tests in the TabCAT battery, this was done to utilise the functionality of the TabCAT system to incorporate language choice for on screen instructions, and participant recorded outcomes and recording of reaction time for both tests (which had not been possible at all for SMT with the previous display only app). This made the use of the MedAvante container forms obsolete so data were no longer collected using this vehicle and when the TabCAT system was re-opened on 12th September 2019 the collection method for FMT and SMT was switched to TabCAT.

When the study was closed in March 2020, all 5 EPAD ENE tests were being collected using TabCAT on a single GCP compliant system with all tests able to be displayed from a single app with instructions in the approved local languages, all data transferred with an upgraded secure synchronisation and feedback process to minimise data loss, with all data available for quality control and correction in an on-line portal, with in-built CTMS and audit trail for all actions in the system. The transformation of TabCAT from an experimental tool to a compliant trial management system has been made possible by the support of EPAD partners and the collaboration of test authors and the team at USCF Memory and Aging Centre.

Processing of 4MT iPad data

Incorrect classification of responses

The original iPad version of the Four Mountains Test records response classification (CORRECT/INCORRECT) and response times in a log file held on the device. Due to a bug the log file records missing responses (i.e., when the subject makes no response after 30s) incorrectly using the most recent active response, however it records the response time correctly (so for items with missing responses the response time is recorded as ≥ 30 s). This means we can identify all missed responses in records extracted from iPad log file data.

Examiners also recorded participant responses (but not response times) on paper records which were then transcribed to the eCRF system. Examiners would be unable to indicate a response to missing trials, therefore these should be correctly classified in the database. These fields thus provide a useful way to double check any corrections we apply to the iPad data.

Action taken: reclassify responses with response times ≥ 30 s as NO_RESPONSE

Longer cut off time

In the original published test and in subsequent implementations (including TabCAT) participants are allowed up to 20s to respond to each item. In the original iPad version, affected by the current bug, a longer period of 30s was allowed. In principle this might mean participants could take longer over responses, but still answer correctly or incorrectly before the cut off. These delayed responses (response time > 20 s and < 30 s) should probably be coded separately so that researchers can construct scores based on the standard 20s time (which would be comparable with TabCAT and earlier published studies). In practice, most responses fall within 20s so this will likely not affect many trials.

Action taken: reclassify CORRECT responses with response times > 20 s and < 30 s as DELAYED_CORRECT and INCORRECT responses with response times > 20 s and < 30 s as DELAYED_INCORRECT.

Researchers would then normally calculate scores based on CORRECT responses only (for comparability with published studies, consistency with TabCAT), while retaining the option to include DELAYED_INCORRECT responses or analyse number of delayed responses.

6.11. 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. 2 records have no information as to whether RBANS was done. 1 record is completely missing RBANS data where the eCRF records RBANS as being done. There are 16 records where the individual did not complete individual RBANS tests and therefore, not all of the indices and totals can be calculated.

6.12. 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.

2 records have no information as to whether "dot_counting", "favourites" and "flanker" were done.

There are 136 "dot_counting" records completely missing where the assessment was performed according to the eCRF.

There are 131 "favourites" records completely missing where the assessment was performed according to the eCRF. In addition, 379 records are completely missing the recognition part of "favourites".

There are 496 "flanker" records completely missing where the assessment was performed according to the eCRF. In addition 2 "flanker" records are partially missing.

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.

FMT was incorporated into the new version of TabCAT from September 2019.

There are therefore, 3 different FMT tables. "four_mountains_edinburgh" contains the older iPad collected data. "four_mountains_medavante" contains MedAvante system data. "four_mountains_tabcat" contains FMT since September 2019.

2 records have no information as to whether FMT was done.

There are 86 records where FMT was performed according to the eCRF but do not have records in any of the 3 FMT tables.

6.13. Four Mountains Task (four_mountains_uedin)

- The Four Mountains Task (FMT)

12 records are partially missing.

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

6.14. Four Mountains Task (four_mountains_medavante)

- The Four Mountains Task (FMT)

117 records are partially missing.

6.15. Four Mountains Task (four_mountains_tabcat)

- The Four Mountains Task (FMT)

1 record is partially missing.

Supermarket Trolley Virtual Reality (SMT)

SMT was incorporated into the new version of TabCAT from September 2019. Before this, SMT data were entered onto the MedAvante system. There are therefore, 2 different SMT tables.

"vr_supermarket_trolley_medavante" contains the MedAvante system data and

"vr_supermarket_trolley_tabcat" contains SMT data since September 2019.

2 records have no information as to whether SMT was done.

There are 92 records where SMT was performed according to the eCRF but do not have records in either of the SMT tables.

6.16. Supermarket Trolley Virtual Reality (vr_supermarket_trolley_medavante)

- Supermarket Trolley Virtual Reality (SMT) individual marks.

The total SMT score can be calculated by counting the number of "Correct" responses from the "st_trial_mark" variables in the data set.

6.17. Supermarket Trolley Virtual Reality (vr_supermarket_trolley_tabcat)

- Supermarket Trolley Virtual Reality (SMT) individual answers and marks

6.18. CDR (cdr)

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

2 records have no information as to whether CDR was done. There are 8 individuals where CDR was performed but have completely missing records.

6.19. MMSE (mmse)

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

2 records have no information as to whether MMSE was done. 1 record has partially missing MMSE data.

6.20. 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. 2 records have no information as to whether A-IADL was done. There are 138 records where A-IADL was performed according to the eCRF but where the total score is missing.

6.21. 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. 2 records have no information as to whether GDS was done. 1 record has partially missing GDS data and 1 record has completely missing GDS data.

6.22. 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. 2 records have no information as to whether PSQI was done. 18 PSQI records are completely missing despite the assessment being performed according to the eCRF. 8 PSQI records have partially missing data.

6.23. 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. 2 records have no information as to whether STAI was done. There are 10 STAI-40 records completely missing where the assessment was performed according to the eCRF. In addition, 13 records are partially missing.

Missing Imaging Data

143 visits have no information as to whether the MRI scan was done. 1 participant is missing "reason_not_performed". It is possible for the MRI scan to have been performed but the imaging measures to be missing. This is often due to the quality of the scan. There are 25 records where the

eCRF records the MRI as being done but no imaging results are available. A further 16 scans have no volumetric imaging results available.

6.24. Imaging Scanner Information (mri_scanner_information)

- Imaging variables giving details of the scanner used for the MRI scan

6.25. Imaging Lacunes and Infarcts (lacunes_infarcts)

- Imaging variables related to lacunes and territorial infarcts.

6.26. Imaging Radiological Read (radiological_read)

- Radiological read imaging variables.

6.27. Imaging Volumes (volumetric)

- Volumetric imaging data.

There are 14 scans with only partial volumetric results available.

IXICO provided the explanation on how to adjust volumes by pseudo total intracranial volume (pTIV):

Here the pTIV is reported as indicator of brain size and usage is recommended if there is evidence for the region being investigated to have a size that scales with the TIV. As such, pTIV is expected to be larger for males than females because is not estimated as in literature (determinant of affine matrix to MNI template), but instead as the magnitude of the affine scaling vector extracted from the transformation matrix. You can use the pTIV as nuisance covariate or as a normalisation factor. In the latter case, conceptually you correct for it by dividing (for example) the whole brain volume by the pTIV (well posed only when QC pass). Normalising in that way, the volume will in principle tend to align with the MNI152 (ver MNI152lin_T1_1mm) volume. Another equivalent way, if you prefer the default TIV approach, would be to “convert” the pTIV into an actual estimate of the subject's TIV by scaling the TIV of the MNI152 (ver MNI152lin_T1_1mm) template brain by the estimated pTIV (for more info [Mazziotta et al., 2001]).

6.28. Dementia Diagnosis (dementia_diag)

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

10 records have no information as to their dementia diagnosis. 1 record gives no date of diagnosis.

1 participant has “type_of_dementia”=MCI. This is incorrectly recorded as a dementia diagnosis in the eCRF.

6.29. HATICE Questionnaire (hatice)

- Healthy Ageing Through Internet Counselling in the Elderly questionnaire.

11 records have no information as to whether the HATICE assessment was performed. A further 38 records are partially missing HATICE as some of the questions were not answered. These data will never be available.

6.30. 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.

6.31. Lifestyle Questionnaire (life)

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

8 records have no information as to whether the lifestyle questionnaire was performed. A further 4 records are partially missing lifestyle information.

6.32. Physical Examination (physical_exam)

- Results from physical examination.

3 records have no information as to whether a physical examination was performed. A further 111 records have no information in "was_ecg_performed".

6.33. Vital Signs (vital_signs)

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

8 records have no information as to whether vital signs were collected. A further 50 records are partially missing vital signs.

6.34. Adverse Events (adverse_events)

- Adverse events information.

6.35. Current Medication (current_medication)

- Current medication information.

6.36. Medical History (medical_history)

- Medical history information.

7. 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
V500.1	Doi:10.34688/epadlcs_v500.1_20.04.29
V.IMI	Doi:10.34688/epadlcs_v.imi_20.10.30