

A4 / LEARN Data FAQ

2025-02-13 (v1.1)

Dataset overview

Q: What are the differences between the 'raw', 'external', and 'derived' datasets?

A: The categories are characterized by their provenance:

Raw	Data entered directly into the coordinating center's electronic data capture (EDC) system.
External	Data provided to the coordinating center from an external provider that pre-processed the data before being transferred. Each external file has a corresponding PDF document that provides background on the dataset.
Derived	Data that was prepared by ATRI biostatisticians that was used to produce the results of the study. These datasets are derived from a combination of the raw and external datasets.

Q: How can I separate the A4 and LEARN cohorts? How can I remove screen-failures from the A4/LEARN datasets?

A: All datasets have a SUBSTUDY field. This field identifies the cohort the participant belongs to:

- A4: Randomized participant of A4 longitudinal study
- LEARN: Enrolled participant of LEARN longitudinal study
- SF: Screen-failed A4 and did not participant in LEARN

Code Guides

Q: How can I reproduce the Table 1 and/or the primary findings of the A4 trial reported in Sperling et al. (2023)?

A: See Intro-to-A4-data.pdf and Intro-to-A4-data.Rmd.

Q: Is there any example code on how to use the various dataset types (i.e. raw, external, and derived)?

A: See a4_learn_data_primer.pdf and a4_learn_data_primer.Rmd for example functions and code.

Rescreens

Q: Are re-screened participants included? How can I identify any participants that were re-screened?

A: Participants who re-screened may appear in the data twice with different BIDs each time. Two fields have been added to the SUBJINFO.csv dataset to identify re-screens:

1. The field 'RESCREENFL' identifies participants that were later re-screened to a new BID.
2. The field 'PREVBID' identifies a previous BID associated with the participant if the participant was previously screened. Mapping any PREVBID to its associated BID is a useful approach in accounting for re-screens. See a4_learn_data_primer.pdf and a4_learn_data_primer.Rmd for example code for this approach on handling re-screens.

Date Fields

Q: The data dictionary indicates the FIELD_FORMAT of this field is a 'date'. Why are all the values in numeric?

A: Each date variable has been replaced with 2 fields: days since consent and days since randomization/baseline (i.e. earliest analysis timepoint or "T0").

“Raw” Data

Coded vs labelled values

Q: Some of the raw datasets have coded values. How can I extract the meta data from data dictionary to find the labelled values of the field?

A: See [a4_learn_data_primer.pdf](#) and [a4_learn_data_primer.Rmd](#) for example functions and code.

Q: For the race category (PTRACE) found in PTDEMOG (Participant Demographics), the different races are coded as 1-6. However, the data set also includes a couple of odd values 1:04, 1:04:05, 1:05, 2:05, and 4:05. In Excel they are reported as times (ex: 1:05:00 AM). How do I interpret these values?

A: The colon-separated strings indicate multiple races were selected. The translation into times is most likely an Excel issue. See [a4_learn_data_primer.pdf](#) and [a4_learn_data_primer.Rmd](#) for example code on converting the variable to distinct binomial fields instead.

Q: Some of the ‘Raw’ data files of the outcomes appear to have all labelled values (e.g. MMSE, CDR). Why is this the case while most other ‘Raw’ data files use coded values?

A: A subset of ‘Raw’ data sets involved making slight modifications (e.g. scoring calculations/adjustments, imputations, and/or additional derived fields) to improve accuracy, usability and completeness of the data. These changes were processed using their labelled values. In order to keep the integrity of the data without re-coding the values back, the labelled values are shared instead for these affected ‘Raw’ data sets:

	CRF Name	CRF Label
1	ADLPQ	ADCS ADL-Prevention Questionnaire for Participant
2	ADLPQSP	ADCS ADL-Prevention Questionnaire for Study Partner
3	CDR	Clinical Dementia Rating
4	CFISP	Cognitive Function Index - Study Partner
5	CFI	Cognitive Function Index - Participant
6	COGDIGIT	Cognitive Test - Digit Symbol Substitution Test
7	COGFCSR16	Cognitive Test - Free and Cued Selective Reminding (16 Pictures)
8	COGLOGIC	Cognitive Test - Logical Memory
9	MMSE	Mini Mental State Exam
10	DOSE	Study Drug Therapy

Safety Assessments

Q: Is adverse event information available?

A: These data were not approved for data sharing at this time. We will release this data in a future date.

Q: Is concurrent medications available?

A: These data were not approved for data sharing at this time. We will release this data in a future date.

Q: Are initial health conditions / comorbidity information available?

A: These data were not approved for data sharing at this time. We will release this data in a future date.

Study Partner Information

Q: Why are there identical RECNO for different BPIDs (study partner ID)? Why do the LEARN study partner information appear to have duplicate records?

A: Particular to the LEARN participants, study partner information was collected twice: once during screening for A4 and again during the LEARN study phase. The A4 and LEARN datasets have been combined in this data share. Therefore, LEARN study partner information likely contains duplicated information across multiple BPIDs.

“External” Data

Q: Where can I find information about the external dataset source and content?

A: Please refer to the [external_datadic.csv](#) and the corresponding PDF file associated with the dataset.

“Derived” Data

Q: Why does the ADQS.csv and SUBJINFO.csv share many of the same fields?

A: ADQS (Analysis Data Questionnaire Scores) contains the scores from all questionnaires (individual and composite scores). The subject information (SUBJINFO) data set was merged into it to allow ADQS to be a complete analysis data set that incorporates all questionnaire outcomes as well as participant-level characteristics. SUBJINFO is still provided separately to serve as a complete participant listing.

Q: Where can I find the longitudinal data for LEARN participants in the COGSTATE_COMPUTERIZED.csv data?

A: COGSTATE_COMPUTERIZED.csv was only prepared for the A4 cohort. Please refer to the external, cogstate-related files (particularly cogstate_battery.csv) for the source data instead which would include all A4 and LEARN data for the computerized Cogstate assessment.

Study visits

Visit information

Q: Where can I find more information about visits and when they occurred?

A: A list of subject visits for each participant can be found in the SV.csv derived data set. The visit_datadic.csv contains the metadata information about each of the visits. Additionally, the ADQS.csv data set has a comprehensive list of assessments and their dates used for the primary analysis. Using the QSADTC field in ADQS.csv may provide additional dates that may be missing from the SV.csv dataset.

Q: Why are the 3 screening visits for LEARN labelled as ‘A4 visit (1,2,3)’ in the visit_datadic.csv?

A: LEARN participants rolled-over from the A4 screening process after the A4 screening visit 3.

Q: Why is ‘LEARN Visit 1 (Baseline: A4 v4-6)’ listed as a visit for the screen-fail (SF) cohort in the visit_datadic.csv?

A: Screen failures in the A4 study can occur anytime between screening visits 1 thru 5. Additionally, if a participant was a rollover to the LEARN study, screen failures in LEARN would have occurred during the baseline visit. Therefore, VISCODEs 001 thru 005 from the A4 study and VISCODE 006 from the LEARN study have been added to the visits data dictionary for the list of potential visits for screen-fails.

VISITGROUP vs EPOCH

Q: Some derived datasets contain an ‘EPOCH’ field. What is the difference between this field and the ‘VISITGROUP’ field found in the visits_datadic.csv file?

A: The “EPOCH” field involves a different method in deriving the values. VISITGROUP is based solely on the VISCODE value, while the EPOCH value incorporates the actual date a study visit occurred. Some examples of differences between the values of these fields include:

- A4 Baseline visits: EPOCH values for A4 baseline visits sometimes may be either defined as a screening visit or a blinded treatment visit based on whether the visit occurred before or after the randomization date. VISITGROUP will always categorize the A4 baseline visit as “Blinded Treatment / LEARN equivalent A4 visit <= 66”.
- Unscheduled visits: EPOCH value may categorize the unscheduled visits to the appropriate longitudinal phase based on when the visit occurred. VISITGROUP does not take into account the study visit date and categorizes unscheduled VISCODEs as “Unscheduled”.
- Missing study visit dates: If the date of the study visit cannot be found, the EPOCH value would be missing. Since VISITGROUP only takes into consideration the VISCODE, the value would align to the appropriate longitudinal phase based on the VISCODE value.

References

Sperling, Reisa A, Michael C Donohue, Rema Raman, Michael S Rafii, Keith Johnson, Colin L Masters, Christopher H van Dyck, et al. 2023. “Trial of Solanezumab in Preclinical Alzheimer’s Disease.” *New England Journal of Medicine* 389 (12): 1096–1107.