

Einstein Aging Study (EAS) Unified Data Documentation

Codebook

Einstein Aging Study Technology and Data Management Core Team

Version: April 17, 2026

Study-first integrated narrative documentation across EAS components

Contents

Part 1. Overview	2
1.1 Introduction	3
1.2 Deliverables	3
1.3 Technical Notes	4
1.4 Key Identifiers & Variables	5
1.5 File Access	6
1.6 Citation & Acknowledgment	6
1.7 Support	7
Part 2. Narrative Codebooks	8
2.1 Blood Biomarkers	9
2.1.1 Alzheimers Disease and Related Dementias	10
2.1.2 Inflammation Biomarkers	11
2.1.3 Oxylipin Biomarkers	13
2.1.4 SNP Genotyping	15
2.2 Clinical Core	17
2.2.1 Clinical Core Protocol	18
2.2.2 Covid Questionnaire	19
2.2.3 Mild Cognitive Impairment (MCI) Diagnoses	20
2.2.4 Stress and Adversity Inventory (STRAIN)	22
2.2.5 Take Home Packet	24
2.3 Sensors	25
2.3.1 Physical Activity and Sedentary Behavior (Sensor: ActivPAL)	26
2.3.2 Air Quality (Sensor: Atmotube Pro)	30
2.3.3 Glucose Monitoring (Sensor: CGM)	32
2.3.4 Sleep Actigraphy and Oximetry (Sensor: Nonin)	34
Part 3. References	42
Part 4. Appendix	45
4.1 SNPs for Genotyping	46
4.2 MCI Diagnoses	53

Part 1. Overview

1.1 Introduction

Narrative Documentation Purpose and Scope

This codebook is provided on behalf of the Einstein Aging Study (EAS) Technology and Data Management (TDM) Core.

The EAS has been going since 1980. We have since 2017, prioritized the use of innovative digital measures for capture of outcomes. This release includes data since 2017 only. To access prior years of data, [Click here to submit a content proposal](#). Over time we will release more data and documentaiton, please refer to future versions of this document as we update and make more data available.

This data release brings all primary releasable datasets current, covering a period from February 2023 through December 2024. With respect to documentation, the EAS provides high level information on our website including codebooks and variables; we also provide the Unified Data Documenation Narrative Codebook for reference. This document consolidates study-level narrative documentation, methods, user notes, and references across EAS data components. Variable-level dictionaries are intentionally excluded because those are maintained in Airtable. Visit the [Einstein Aging Study website](#) for more information, and [Click here to view publications](#).

Note: This document consolidates study-level narrative documentation. Variable-level dictionaries are maintained in Airtable, and additional high-level information can be found on the study website.

Study Context and Program Overview

The broader EAS study began in earnest in 1993, with additional data from participants as early as 1982 from two earlier studies that merged to become part of EAS. Since 1993, there have been various iterations of an NIH-funded program project grant, which has ensured some continuity in variables and participants but has also resulted in several phases of the study with differences in protocol and measurements. Since its inception, EAS has been multidisciplinary; early details and protocols are described elsewhere. More detailed information can be found on the [Einstein Aging Study website](#).

Recruitment was completed via systematic probability sampling of the NY Voter Registration List for Bronx County To date, recruitment follows a racial/ethnic distribution comparable to that of the US Census data for Bronx County. Compared to the U.S. Census data, we had higher representation of Black/African Americans (44.8% vs 22.0%) and Hispanic/Latino (40.7% vs 32.9%) adults. Overall Population in Bronx County, New York is 1.4 million (12.9% 65+). Data extracted from the American Community Survey 2022, 5-year Detailed Table. Data extracted via the R package, tidycensus (<https://walker-data.com/tidycensus>), with variables selected from <https://api.census.gov/data/2022/acs/acs5/variables.html>.

Funding Acknowledgment

This research was supported by grant U24AG092760 from the National Institute on Aging, part of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

1.2 Deliverables

Release Deliverables

This data release includes the following core documentation and data deliverables intended to support use and interpretation of the EAS datasets:

- EAS Unified Data Documentation Narrative Codebook
- Data Dictionary in Airtable

- The Data, provided in flat file format
- Reports & Data Quality Control Files

Airtable Data Dictionary

Measure and variable-level information is provided in the Airtable Data Dictionary, and links to measures and variables can be found below.

- [EAS Data Dictionary Variable View \(Airtable\)](#)
- [EAS Data Dictionary Measure View \(Airtable\)](#)

Notable Updates

The deliverables of this data release were created to mirror prior data releases, with respect to data granularity and structure. There are subtle differences in process, so you will notice renamed columns, standardization of columns to be of similar case (i.e., snake_case), and other details, abbreviated below. Due to changes in EAS data management policies, additional data not contained herein is available upon request with approved concept proposal.

Flat File Naming Conventions

An example of deliverable flat file naming convention found in the release is as follows:

- EMA surveys (all), session-level:
`tidy_eas_ema_all_surveys_session_level_{ts_fn}.csv`

Note: {ts_fn} is a placeholder for the filename-friendly Sys.time() value at the time of exporting files.

1.3 Technical Notes

About this Data Release

This data release includes all data from **May 2017 through December 2025** for all participants completing the new protocol. If you are amid a revise and resubmit with a prior dataset, do not use the latest.

Note: If these files are stacked with prior releases, we did switch mobile apps for data collection, and care should be paid to how this merge should take place. Please contact with Dr. Nelson Roque (nur375@psu.edu) or Sarah Logan (szl6413@psu.edu) of the TDM Core for more information.

Data Processing and Reproducibility

This data release was fully prepared in R in support of the TDM Core's mission for open and reproducible science. Processing is completed via a set of R functions, which are formalized as the R package pipes. [Click here to access the tidypipes GitHub repository.](#)

Quality Control

The following are the included quality control reports, provided for all data users.

REPORT_code_{ts_fn}.csv	REPORT_wc_acq_leisure_{ts_fn}.csv
REPORT_env_{ts_fn}.csv	REPORT_wc_all_instruments_{ts_fn}.csv
REPORT_fns_table_{ts_fn}.csv	REPORT_wc_discrimination_{ts_fn}.csv
REPORT_missing_cdr_{ts_fn}.csv	REPORT_wc_emotional_health_{ts_fn}.csv
REPORT_missing_day1_date_a_{ts_fn}.csv	REPORT_wc_fatigue_{ts_fn}.csv
REPORT_missing_day1_date_b_{ts_fn}.csv	REPORT_wc_lifesatisfaction_{ts_fn}.csv
REPORT_missing_day2_date_{ts_fn}.csv	REPORT_wc_loneliness_{ts_fn}.csv
REPORT_missing_day3_date_{ts_fn}.csv	REPORT_wc_neighborhood_{ts_fn}.csv
REPORT_missing_gcrc_{ts_fn}.csv	REPORT_wc_panas_{ts_fn}.csv
REPORT_packages_{ts_fn}.csv	REPORT_wc_personality_{ts_fn}.csv
REPORT_track_both_0_and_99_{ts_fn}.csv	REPORT_wc_physicalfunctioning_{ts_fn}.csv
REPORT_track_check_99_before_0_{ts_fn}.csv	REPORT_wc_promisSatisfaction_{ts_fn}.csv
REPORT_wc_reap_{ts_fn}.csv	REPORT_wc_socialactivity_{ts_fn}.csv
REPORT_wc_subjectivestress_{ts_fn}.csv	REPORT_wc_support_{ts_fn}.csv
REPORT_wc_trackform_{ts_fn}.csv	REPORT_wc_trackform_meta_{ts_fn}.csv
REPORT_wristband_ids_{ts_fn}.csv	

1.4 Key Identifiers & Variables

Identifiers

Participants across prior releases had various identifiers (e.g., Id, ID, Subject Id, EAS ID, GCRC Id). In an effort to reduce confusion, all participants are now labeled with `textttparticipant_id` that is standard across deliverables.

Timepoint Identifiers

Across EAS data there are multiple variables that track when participants completed various study components. The following defines each Timepoint Identifier you may encounter in the data.

- **burst**: An integer representing a given participant's number of visits to the Einstein clinic since 2017. Burst = 0 indicates that the participant never did the EMA. Burst = 99 means that the participant did the EMA in previous waves but did not do it that wave. This can change from year to year.
- **burst_cumulative**: An integer representing number of visits to the einstein clinic since 2017. (Does not include Burst 0 or 99)
- **wave**: An integer representing number of visits to the einstein clinic over time.
- **device_date**: The calendar date in YYYY/MM/DD format for a given sensor

Cohort Definition

Given that the EAS has existed and been recruiting for multiple years, there is a binary cohort variable that allows researchers to sort participants by cohort. Cohort is distinguishable as follows:

- If `participant_id` \geq 13000 then `cohort=2`; new cohort, 2023 and beyond
- If `participant_id` \geq 12000 then `cohort=1`; new cohort, 2017 to early 2023
- If `participant_id` $<$ 12000 then `cohort=0`; old cohort, prior to 2017

Vintages

We as a team decided to make protocol changes from the 2017 grant to now - as a result, we have two vintages of data: The 2017–2022 vintage, and the 2023–2028 vintage.

A non-exhaustive example of changes you can expect to see in the 2023–2028 vintage include added take-home-packet variables, added sensor variables, and the addition of Atmotube, ActivPAL and CGM sensor data streams.

Dates & Times

In an effort to clarify time stamps across all data, we made some standardization decisions.

In the EMA data, we standardize in the following way: all date/timestamps were reduced to a set of columns, herein: `dt_start_date`, `dt_start_time`, `dt_end_date`, `dt_end_time`.

In sensor data, we have standardized all devices to have a `device_date`

We include the three visits to the clinic as the variables `day_1_appointment_date`, `day_2_appointment_date`, and `day_3_appointment_date`.

1.5 File Access

General File Access

File access is available for all Einstein users via Dropbox. For all other users, see institutional agreement.

Notes for R and SAS Users

In the December 2022, it was noted that importing CSVs into SAS presented a couple situations to resolve: (1) long column names; (2) missing data codes. In this release, you will find every file existing two times, files suffixed with `_R_READY.csv` and `_SAS_READY.csv` to indicate which files are prepared according to the program it is intended to be used with.

1.6 Citation & Acknowledgment

Researchers must acknowledge use of study data with the following language:

The Einstein Aging Study (EAS) has grant support from the National Institutes of Health National Institute on Aging Grant AG003949. The content of this (article, paper, abstract, etc) is solely the responsibility of the authors and does not necessarily represent the official views of the NIA or the NIH. We thank the study staff and all of the participants who have enrolled in the EAS. Data were from the Einstein Aging Study (EAS): Study information, codebooks, and a form to request data access are available through the Einstein Aging Study website, <https://einsteinagingstudy.com>

If Actigraphy, Nonin, or Ambient Light (Sleep Project) data are used, please also acknowledge the following::

Actigraphy, Nonin, and Ambient Light data from The Einstein Aging Study (EAS) has grant support from the National Institutes of Health National Institute on Aging Grant AG003949, and if sleep data collected prior to 2023 are used - R01AG062622. The content of this (article, paper, abstract, etc) is solely the responsibility of the authors and does not necessarily represent the official views of the NIA or the NIH. We thank the study staff and all of the participants who have enrolled in the EAS. Data were from the Einstein Aging Study (EAS): Study information, codebooks, and a form to request data access are available through the Einstein Aging Study website, <https://einsteinagingstudy.com>

If Ecological Momentary Assessment (EMA) data, including survey or M2C2 data are used, please additionally acknowledge the following:

The Mobile Monitoring of Cognitive Change (M2C2) assessments were developed as part of the National Institute on Aging (NIA) Grant U2CAG060408 and Open Measures Network Initiative for Alzheimer's Disease and Related Dementias (OMNI ADRD) NIA Grant U24AG092760. Please acknowledge these grants in your presentations and publications, as this recognition helps us track and report to the NIA about our dissemination activities and their impact. Your support in this regard is greatly appreciated! Please use the following acknowledgement for presentations and manuscripts: "This work utilized Mobile Monitoring of Cognitive Change (M2C2) assessments developed under NIA Grant U2CAG060408 and NIA Grant U24AG092760.

1.7 Support

Note: Encounter an issue? Reach out to the Tech and Data Management Core by emailing Dr. Nelson Roque (nur375@psu.edu), Sarah Logan (szl6413@psu.edu), and Nilietta Bravo (neb5486@psu.edu).

Something Missing?

The Einstein Aging Study has been going since 1980. This release only incorporates data from 2017 onwards. If you are looking for data from a prior year, [Click here to submit a content proposal](#), and reach out to Mindy Katz (mindy.katz@einsteinmed.edu) and Nelson Roque (nur375@psu.edu).

If you notice an expected deliverable is missing, please contact Dr. Nelson Roque (nur375@psu.edu), Sarah Logan (szl6413@psu.edu), and Nilietta Bravo (neb5486@psu.edu). In your email, please make sure to note the filename of any prior versions of the dataset expected. Please note that based on existing DUAs and Reliance agreements, only de-identified data is being shared in this release. If your work for your grant aims, or concept proposal require anything beyond de-identified data, please contact Dr. Nelson Roque to discuss furnishing of additional data, and if amendments to existing agreements are required.

Technical Issues?

If you encounter any issues working with this dataset (e.g., due to volume, naming conventions, etc), reach out! The Tech and Data Management Core is happy to help troubleshoot and provide support.

Suggestions? Feedback? Want to Donate Code?

We're committed to improving our process. Please reach out with any feedback or suggestions you have. The TDM Core also accepts and welcomes donated code (e.g., if you have code for creating composites that you use in your lab, or feedback on additional flags to add to the data).

Part 2. Narrative Codebooks

2.1 Blood Biomarkers

The blood biomarker protocol collects and analyzes blood samples to characterize biological processes related to aging, health, and risk for age-related diseases, including Alzheimer's disease and related dementias (ADRD). Blood samples are processed using a range of laboratory assays that measure biomarkers reflecting multiple physiological domains, such as neurodegeneration, inflammation, metabolic function, and lipid signaling pathways.

This protocol includes both established clinical markers and emerging research biomarkers, including state-of-the-art assays for Alzheimer's disease pathology (e.g., phosphorylated tau and amyloid-related markers). By capturing information across multiple biological systems, these measures enable researchers to investigate how molecular processes relate to cognitive function, environmental exposures, lifestyle factors, and disease risk.

2.1.1 Alzheimers Disease and Related Dementias

	Snapshot
Project/Core	Biomarker Core
Funding period / Vintages	2017–2022; 2023–2028

Filenames in Release

The following file(s) are included in release:

`BIO_adrd_{ts_fn}.csv`

Overview

These data include blood-based biomarkers associated with Alzheimer’s disease and related dementias (ADRD), measured using state-of-the-art assays (e.g., phosphorylated tau [p-tau] and amyloid-beta [$A\beta$]). These markers reflect pathological processes such as amyloid deposition and tau-related neurodegeneration and are widely used to study biological aging, cognitive decline, and dementia risk. Researchers may use these measures to examine relationships between molecular pathology, cognitive performance, environmental exposures, or other health outcomes.

Notes on Equations for Glomerular Filtration Rate (GFR) Variables

Estimating equations for Glomerular Filtration Rate (eGFR) can be found in the study by Inker et al., titled New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race (2021). [Click here to read the NEJM study on eGFR](#)

Using `egfr_crAS` and `egfr_crASR` in analytic models has different trade-offs:

- `egfr_crASR` tends to overestimate measured GFR in Black participants and slightly overestimates it in non-Black participants.
- `egfr_crAS` underestimates measured GFR in Black participants and overestimates it in non-Black participants.
- `egfr_crAS` has a larger differential bias compared to `egfr_crASR`.
- Using `egfr_crASR` may overcorrect for race differences in biomarker levels, since race is used in the model twice.

Data Cleaning and Analysis

This section summarizes outliers that were removed from the dataset:

- One NFL observation was removed because it was 10 standard deviations (SD) away from the mean value across the entire sample.
- Two pTau181 observations were removed because they were 10 SD away from the mean value across the entire sample.
- There are 41 missing values for eGFR. These are true missing values (i.e., no blood was drawn).
- pTau181 can be log-transformed prior to analysis to yield a less right-skewed distribution.

2.1.2 Inflammation Biomarkers

Snapshot	
Project/Core	Biomarker Core
Funding period / Vintages	2017–2022; 2023–2028

Filenames in Release

The following file(s) are included in release:

BIO_infl_{ts_fn}.csv

Overview

Inflammatory biomarkers measure immune system activity and systemic inflammatory responses. Chronic inflammation has been implicated in many age-related diseases, including cardiovascular disease and neurodegeneration. These markers enable researchers to examine relationships between immune activation, environmental exposures, aging processes, and cognitive health.

Description

These variables represent biomarkers from blood plasma that pertain to inflammation. Blood was collected at the start (pre-EMA)¹ and conclusion (post-EMA) of each EMA burst, which are named “Day 2” and “Day 3” (respectively) in EAS nomenclature. The pre-EMA (“Day 2”) blood sample came from a non-fasting blood draw, and the post-EMA (“Day 3”) blood sample came from a 12 hour fasting blood draw. A certified phlebotomist collected all samples (both pre- and post-EMA) between 7 AM and 11 AM at the Albert Einstein College of Medicine. Blood (5 mL) was collected in EDTA collection tubes² via venous puncture to assess basal and stimulated cytokine levels and C-reactive protein (CRP). Other blood aliquots were banked for protein quantification of additional analytes (e.g., hormones [see below]).

To determine basal inflammation and CRP, whole blood was centrifuged at 1500g for 15 min at room temperature. The supernatant was aliquoted and stored at -80°C .

To determine stimulated cytokines, whole blood (1 mL) was exposed *ex vivo* to 1 $\mu\text{g}/\text{mL}$ of bacterial lipopolysaccharide (LPS; E. coli 055: B5, Sigma-100 mg)³ on a rotational shaker at 37°C in 5% CO_2 for 2 hours. Samples were then centrifuged at 1500g for 15 min at room temperature. The supernatant was aliquoted and stored at -80°C .

Stimulated cytokines are preceded with an “s_” to differentiate them from the non-stimulated basal cytokines.

Cytokines and C-reactive protein (CRP) were measured via Meso Scale Diagnostics (MSD; Rockville MD) multiplex arrays from plasma.

In addition, basal and stimulated cytokine composite measures were calculated for each participant for each wave (i.e., averaging across pre- and post-EMA measurements; see “Associated Papers” for more details on composites). Missing data are denoted as “N/A”. The minimum detection limit for cytokine and CRP was reported in the kit product inserts with the following values:

Name	Detection Limit
IL1b	0.05 pg/mL
IFN γ ⁴	0.37 pg/mL
IL4 ⁵	0.02 pg/mL
IL6	0.06 pg/mL
IL8	0.07 pg/mL
IL10	0.04 pg/mL
TNF α	0.04 pg/mL
MIF ⁴	4.3 pg/mL
CRP	1.33 pg/mL

All samples were run in duplicate. Sample pairs with coefficients of variation (CVs) greater than 15% were rerun when possible. Confirmed values below the minimum detection limit were replaced with zeros.

¹ The blood draw at the start of the burst was incorporated on 8/15/2017. Participants enrolled prior to this date will not have the pre-EMA (“Day 2”) inflammatory data for burst 1.

² The LPS-stimulated collection tubes were switched from Sodium Heparin tubes to EDTA tubes after February 6, 2018 (one additional participant’s LPS-stimulated sample was collected in heparin tubes on February 22, 2018). LPS stimulated samples collected before this date have been excluded. The samples collected in sodium heparin tubes are denoted with a 1 in the “HEPLPS” column. Unless otherwise stated, stimulated samples were collected in ethylenediaminetetraacetic acid (EDTA).

³ Lipopolysaccharide (LPS) is an antigen (i.e., the Lipid A component of gram-negative bacteria cell wall). We use the *ex vivo* stimulation paradigm to mimic an immune challenge, as LPS stimulates/activates immune cells in whole blood. From this we can quantify the inflammatory response that cells in whole blood generate to an immune challenge.

⁴ Interferon gamma (IFN γ) and macrophage migration inhibitory factor (MIF) were added for the 2023+ samples.

⁵ Interleukin 4 (IL4) was phased out of being measured during the study due to the majority of samples not being able to be measured for this analyte as the concentration was below the limit of detection.

2.1.3 Oxylipin Biomarkers

	Snapshot
Project/Core	Biomarker Core
Funding period / Vintages	2017–2022; 2023–2028

Filenames in Release

The following file(s) are included in release:

BI0_oxy_{ts_fn}.csv

BI0_oxy_lod_{ts_fn}.csv

Overview

Oxylipins are bioactive lipid mediators derived from polyunsaturated fatty acids that regulate inflammation, vascular function, and immune signaling. Oxylipin profiles can provide insight into metabolic pathways involved in inflammatory regulation and may be used to study links between lipid metabolism, environmental exposures, and disease-related processes.

Description

These variables represent oxylipin biomarkers from blood plasma. Blood was collected at the start (pre-EMA)¹¹ and conclusion (post-EMA) of each EMA burst, which are named “Day 2” and “Day 3” (respectively) in EAS nomenclature. The pre-EMA (“Day 2”) blood sample came from a non-fasting blood draw, and the post-EMA (“Day 3”) blood sample came from a 12-hour fasting blood draw. A certified phlebotomist collected all samples (both pre- and post-EMA) between 7 AM and 11 AM at the Albert Einstein College of Medicine. Blood (5 mL) was collected in EDTA collection tubes²² via venous puncture to assess basal and stimulated cytokine levels and C-reactive protein (CRP). Other blood aliquots were banked for protein quantification of additional analytes. All oxylipin were assayed using the fasting blood sample collected from “Day 3”.

GCRC Processing Procedures

EAS plasma samples were thawed on ice and extracted using a modified Smedes protocol¹. 100 μ L of plasma was combined with 5ul of antioxidant BHT/EDTA (0.2ng/mL) and 20ul of 1000nM surrogate containing 9-HODE-d4, 9(10)-EpOME-d4, 12-HETE-d8, 14(15)-EpETrE-d11, 17-HDoHE-d5 and 9(10)-DiHOME-d8. Liquid-liquid extraction was performed to isolate the fatty acids and oxylipins from the plasma. Samples were divided for extraction of total (esterified + nonesterified) and nonesterified oxylipins. Hydrolysis of oxylipins was performed for the extraction of total oxylipins, where the samples were hydrolyzed using 0.5 M sodium methoxide in methanol. The extraction of the nonesterified oxylipins did not involve the hydrolysis step. After, samples were subjected to SPE extraction using 3 mL Chromabond HLB columns (Machery-Nagel: Duren, Germany). The collected samples were then dried down and reconstituted in 1:1 methanol:acetonitrile mixture containing 100nM of 1-cyclohexyl ureido, 3-dodecanoic acid (CUDA) as an internal standard for intrasample validation.

Oxylipin measurement was carried out using a liquid chromatograph-mass spectrometer (Waters Xevo TQD, Acquity I-Class; MA, USA) and separated using a CORTECS UPLC C18 2.1 x 100 mm with 1.6 μ M particle size column

¹The blood draw at the start of the burst was incorporated on 8/15/2017. Participants enrolled prior to this date will not have the pre-EMA (“Day 2”) inflammatory data for burst 1.

²The LPS-stimulated collection tubes were switched from Sodium Heparin tubes to EDTA tubes after February 6, 2018 (one additional participant’s LPS-stimulated sample was collected in heparin tubes on February 22, 2018). For details, refer to the inflammation variable codebook and documentation.

(Waters; MA, USA). Column temperature was set to 40°C, sample injection volume 5ul, and a flow rate of 0.5 mL/min. Solvent A was water with 0.1% acetic acid and solvent B was acetonitrile:isopropanol 90:10.

Mass spectrometry analysis was carried out using negative electron spray ionization with a capillary voltage set at 2.0 KV, source temperature 150°C, desolvation flow of 1000 L/hr, and desolvation temperature of 600°C. Standards for each oxylipin were previously infused to optimize parameters for cone voltage, collision energy, and analysis in multiple reaction monitoring of parent and daughter ion molar masses. Calibration curves for each oxylipin were measured and quantified using the 5 dilution standard samples. The LOQ and LOD of each oxylipin was calculated using the standard deviation of the response and slope of the calibration curves ($LOQ=10\sigma/S$; $LOD=3.3\sigma/S$). All data was processed using TargetLynx software (Waters; MA, USA).

EAS Data User Notes

There are N=306 Burst 1 EAS participants included in this data set.

Generating enzymes are the most commonly attributed enzymes at the production of this code book. Some may have multiple enzymes involved or are currently undetermined.

Limit of Detection (LoD), not limit of Quantification (LoQ), is recommended to be used as the cut-off points for data imputation procedures before data analyses. Before performing any transformation to the data or any analyses, LoD value should be applied to each variable in the Total fraction dataset to impute any values below LoD with the LoD value.

2.1.4 SNP Genotyping

<hr/>	
Snapshot	
<hr/>	
Project/Core	Biomarker Core
Funding period / Vintages	2017-2022; pending for 2023+

Filenames in Release

The following file(s) are included in release:

```
BI0_genetics_apoe{ts_fn}.csv
```

Overview

Single nucleotide polymorphism (SNP) genotyping data identify genetic variants across the genome. These variants (e.g., APOE4) can be used to examine genetic risk factors for disease, perform candidate gene or genome-wide analyses, and construct polygenic risk scores related to traits such as cognitive decline, metabolic function, or disease susceptibility.

EAS SNPs for Genotyping

The GWAS genotyping was performed on 784 subjects using Illumina Global Screening Array-24 v3.0 Kit, on which SNP rs7412 was directly genotyped; rs429358 was imputed ($R_{sq}=0.95$) on Michigan Imputation Server 2 based on 1000 Genome Panel Phase 3 panel.

For the full list of SNPs for Genotyping, please see the appendix [4.1](#)

Secondary list for PD related genes: chosen as significant SNPs from previous GWAS and were replicated in one or the other of the studies listed below.

SYT11: rs34372695

ACMSD: rs10928513

FGF20: rs591323

CCDC62/HIP1R: rs12817488

STX1B: rs4889603

NMD3: rs34016896

SREBF1/RAI1: rs11868035

APOE Genotypes

This document outlines the details and structure of the datasets generated for ApoE genotype data, including dataset composition, methodology, and key notes for interpretation.

Two primary datasets were created:

- **ApoE Genotype New EAS.xlsx**
 - Contains data for participants in the new EAS master dataset (December 2022 version).
 - Sample size: $n = 306$ participants with ApoE genotype information.
 - E4 carrier rate: 26%.
- **ApoE Genotype All EAS.xlsx**

- Contains data for all EAS subjects with ApoE genotype information.
- Sample size: $n = 1600$.
- E4 carrier rate: 24%.
- **Frequency of ApoE Genotype and E4 Carrier in New and All EAS.xlsx**
 - Includes frequency tables for:
 - ◇ ApoE genotypes
 - ◇ ApoE E4 carriers in both datasets (new EAS and all EAS)

Methodology

- **Combining Sources**
 - ApoE genotype data were derived from two sources: the original genotype source and the GWAS source.
 - Coding discrepancies between sources (e.g., “e3/e2” vs. “e2/e3”) were reconciled.
 - For discrepant cases, values from the original source were used, as recommended by Kenny.
 - A list of discrepant records is provided separately for reference.
- **Subsets Created**
 - New EAS Dataset: Includes participants from the December 2022 EAS master dataset with available ApoE genotype information.
 - All EAS Dataset: Includes all EAS subjects with ApoE genotype information.

2.2 Clinical Core

The Clinical Core is a standardized assessment battery designed to capture key clinical, cognitive, and health-related measures relevant to aging and dementia research. The battery is largely derived from the Uniform Data Set (UDS) developed by the National Alzheimer's Coordinating Center (NACC), which provides a common framework used across many Alzheimer's disease research studies.

In addition to replicating core UDS components, the Clinical Core includes several enhancements and additional measures that expand the scope of assessment to capture broader aspects of health, lifestyle, and cognitive functioning. These additions allow researchers to examine a wider range of factors related to aging, cognition, and dementia risk while maintaining compatibility with established ADRD research protocols.

2.2.1 Clinical Core Protocol

	Snapshot
Project/Core	Clinical Core
Funding period / Vintages	2017–2022; 2023–2028

Filenames in Release

The following file(s) are included in release:

CLINIC_summary_file_ts_fn.csv

Overview

collects information on participants' health status, daily habits, cognitive concerns, mood, and lifestyle factors. These measures provide contextual information about participants' physical and mental health and can be used to characterize the study population or examine relationships between health behaviors, mood, and cognitive outcomes.

The Clinical Core codebook is included in this unified documentation set and includes the following measures:

Benson Complex Figure	Medical History
Benson Complex Figure – Recall (Delayed)	Medications
Benson Complex Figure – Recognition (Delayed)	MINT (Multilingual Naming Test)
Blessed	Montreal Cognitive Assessment
Category Fluency	MOS Sleep Scale
Clinical Core	Number Span Test: Backward
Cognitive Change Index (CCI)	Number Span Test: Forward
Craft Story 21 Recall (Delayed)	Pain Questionnaires
Craft Story 21 Recall (Immediate)	Smoking and Alcoholic History
Demographics	Social Activities
Diagnosis and Rating Variables	Social Network: Social and Caregiver Strain
FAST	Trail Making Test
Family Medical History	Verbal Fluency: Phonemic Test
Filter Variables	WAIS-III Block Design
Free and Cued Selective Reminding Test (Delayed)	WAIS-III Digit Symbol
Free and Cued Selective Reminding Test (Immediate)	Wechsler Memory Scale: RLM (Delayed)
GDS	Wechsler Memory Scale: RLM (Immediate)
Identification Variables	WRAT-4 Reading Subtest
Lawton Brody	Life Events

Note: Variable-level tables are intentionally excluded from this narrative document and are managed in [Airtable](#).

Ambulatory Cognitive Assessment

In collaboration with M2C2, we have developed additional ambulatory cognitive assessment scores for each of the ambulatory cognitive assessments, available upon approved concept proposal. Primary metrics are available as part of this data release.

2.2.2 Covid Questionnaire

	Snapshot
Project/Core	Clinical Core
Funding period / Vintages	2017–2022

Filenames in Release

The following file(s) are included in release:

```
CLINIC_covid_baseline_both_{ts_fn}.csv  
CLINIC_covid_fu_both_{ts_fn}.csv  
CLINIC_covid_vaccine_both_{ts_fn}.csv
```

Overview

gathers information about participants' experiences during the COVID-19 pandemic, including infection history, symptoms, and potential impacts on health and daily functioning. These data allow researchers to examine how COVID-19 exposure and pandemic-related disruptions may influence physical health, cognitive function, and well-being.

COVID questionnaire documentation is included as a dedicated component within the unified documentation set. This questionnaire includes the following:

- Capture symptom, testing, diagnosis, vaccination, and recent-change check-in context.
- Preserve wording and administration context from the original source instrument documentation.
- Maintain variable-level coding dictionaries and detailed tables externally in Airtable.

2.2.3 Mild Cognitive Impairment (MCI) Diagnoses

	Snapshot
Project/Core	Clinical Core
Funding period / Vintages	2017-2022; 2023-2028

Filenames in Release

The following file(s) are included in release:

CLINIC_summary_file_ts_fn.csv

Overview

diagnosis data indicate whether a participant meets clinical criteria for MCI, a condition characterized by measurable declines in memory or other cognitive abilities that exceed typical age-related changes but do not meet criteria for dementia. These data can be used to classify participants by cognitive status and to study risk factors and progression toward dementia.

For this data release, four classifications of MCI are provided, each with their own assumptions and data inputs. Review the notes below to determine the best fit for your analyses. Please reach out to Mindy Katz (mindy.katz@einsteinmed.edu) for further clarification on classifications.

All variables below are binary coded (0 = No, 1 = Yes).

- **mci**: Jak Bondi classification with age, sex, and education corrections. Does not include nonrandom missing data information.
- **mci1**: Jak Bondi classification with age, sex, and education corrections. Includes nonrandom missing data information.
- **j_mci_race_adj**: Jak Bondi classification with age, sex, education, and race corrections (race compares White vs. Black; Hispanic group too small). Does not include nonrandom missing data information.
- **j_mci_race_adj1**: Jak Bondi classification with age, sex, education, and race corrections (race compares White vs. Black; Hispanic group too small). Includes nonrandom missing data information.
- **j_mci_cog**: Cognitive impairment status in Jak Bondi definition using TRA1 and TRB1
- **j_mci**: Jak Bondi MCI status determined based on cognitive impairment status using TRA1 and TRB1 and considering ADL function
- **jmci_cog**: Cognitive impairment status in Jak Bondi definition using TRAILA and TRAILB
- **jmci**: Jak Bondi MCI status determined based on cognitive impairment status using TRAILA and TRAILB and considering ADL function
- **p_mci**: Peterson MCI status using TRA1 and TRB1
- **p_amci**: amnestic Peterson MCI using TRA1 and TRB1
- **p_namci**: non-amnestic Peterson MCI using TRA1 and TRB1
- **p_namci_subtype**: amnestic Peterson MCI subtypes using TRA1 and TRB1
- **p_amci_subtype**: non-amnestic Peterson MCI subtypes using TRA1 and TRB1
- **pmci**: Peterson MCI status using TRAILA and TRAILB

- `a_mci`: amnesic Peterson MCI using TRAILA and TRAILB
- `na_mci`: non-amnesic Peterson MCI using TRAILA and TRAILB
- `na_mci_subtype`: amnesic Peterson MCI subtypes using TRAILA and TRAILB
- `a_mci_subtype`: non-amnesic Peterson MCI subtypes using TRAILA and TRAILB

For the full list of MCI Diagnoses, please see the appendix [4.2](#)

2.2.4 Stress and Adversity Inventory (STRAIN)

	Snapshot
Project/Core	Clinical Core
Funding period / Vintages	2017–2022; 2023–2028

Filenames in Release

The following file(s) are included in release:

CLINIC_strain_ts_fn.csv

Overview

The Stress and Adversity Inventory (STRAIN) is a structured questionnaire that assesses exposure to acute and chronic stressors across the lifespan. It captures information about the timing, severity, and types of stressful experiences. STRAIN data enable researchers to examine how lifetime stress exposure relates to health, aging, cognitive functioning, and disease risk.

“The STRAIN, or Stress and Adversity Inventory, is a NIMH/RDoC-recommended instrument that efficiently and reliably assesses a person’s cumulative exposure to stress over the life course. The measure is entirely online and systematically inquires about a diverse array of acute life events (e.g., deaths of relatives, job losses, negative health events) and chronic difficulties (e.g., ongoing health problems, work problems, relationship problems, financial problems, etc.) that have implications for human health and well-being. Stressors occurring in early life (e.g., childhood maltreatment or neglect, parental loss/separation, etc.) are also queried in detail. Respondents are asked to rate the severity, frequency, timing, and duration of each stressor they endorse. Questions that are inappropriate (based on a participant’s demographic characteristics) are automatically omitted from the interview (e.g., female reproductive health questions for male participants, questions about children for persons without children). The instrument can be self-administered by users at a computer or can be administered by an interviewer who follows a series of simple on-screen prompts. Because the STRAIN is embedded in an automated, online interviewing environment, the interview can be completed almost anywhere, including in the clinic, research laboratory, or classroom. Presently, we have an adolescent version of the STRAIN (Adolescent STRAIN) that is available in English, and an adult version of the STRAIN (Adult STRAIN) that is available in English, Spanish, German, Swiss (High) German, Brazilian Portuguese, Croatian, and Italian (to begin using the STRAIN, complete the STRAIN Setup Form).

The average time needed to complete the STRAIN is 25 minutes, with a range of approximately 18-30 minutes based on the population being interviewed. Because there are multiple follow-up questions for each endorsed stressor (i.e., that assess severity, frequency, timing, and duration), there are approximately 220 questions that can be asked in all. Based on this information, the system produces 455 variables that are used to assess an individual’s cumulative exposure to stress over the life course. Using this raw data, we can presently create more than 115 different cumulative life stress summary variables and life charts that summarize a person’s lifetime stress exposure. Analyses can in turn be based on a number of factors, including stressor severity and/or the timing of stress exposure (e.g., Early Adversity vs. Distant vs. Recent Life Stress). More sophisticated analyses can be performed by focusing on stressors occurring in particular life domains (e.g., Housing, Education, Work, Health, Marital/Partner) or that have particular core characteristics (e.g., Interpersonal Loss, Physical Danger, Humiliation, Entrapment, Role Change).

Several other interview-based measures have been developed for assessing life stress over relatively short periods of time (e.g., a few months or years). The STRAIN is not a substitute for these systems, but rather is an alternative that can be used when the goal is to quickly and efficiently collect information about stressors occurring over the lifespan as opposed to over a few months or years. The STRAIN accomplishes this goal by combining the sophistication of an interview-based measure of life stress with the simplicity of a self-report instrument.”

For more information and associated publications: [Click here to read more about STRAIN](#)

Incorporating the STRAIN into the EAS Protocol

Funding to support adding the STRAIN instrument to EAS began in July 2018. STRAIN was implemented in EAS around **Sept. 28, 2018**; participants with an EAS baseline date before this would not have been given STRAIN.

Data Notes

Participant Notes

These are participant-level data, including notes regarding the missingness of STRAIN and EMA data (or broader notes such as overall loss of contact with participants precluding further data collection). These notes are a combination of reports from Einstein researchers who work with participants on-location to administer the STRAIN and notes from EMA data managers.

STRAIN Variables

The primary STRAIN datasets include variables described in the source codebook. STRAIN is administered online; Dr. George Slavich's team at UCLA collects and manages raw STRAIN-item data and shares participant-level summary datasets with EAS upon request. Item-level data is generally NOT included by default, and complete survey responses are typically prioritized, though partial completion files can be requested.

'All Data'

This dataset includes STRAIN variable data for participants who fully completed the STRAIN survey, and each participant may have more than one entry if they completed the STRAIN more than once. There are a number of flag variables (described in the codebook) for whether this data is from the first administration of the STRAIN or subsequent administrations. This data also includes participant notes from 2021, so some IDs in the dataset may not have STRAIN data.

Note: Participants are meant to take the STRAIN once at baseline, but occasionally some have completed it multiple times and/or may have partial completions, which may result from purposeful or accidental partial completions of the survey.

'Incomplete'

This dataset contains participant-level information for partial or incomplete STRAIN administrations. Participants may have multiple entries in this dataset if they have multiple incomplete STRAIN surveys.

Note: There are multiple digital pages of questions for the STRAIN, with the last page being 408.

'No Duplicates 1st STRAIN'

This dataset includes STRAIN variable data for participants who fully completed the STRAIN survey, and each participant has a single entry corresponding to data from their FIRST STRAIN administration. This data also includes participant notes from 2021, so some IDs in the dataset may not have STRAIN data.

Note: participants are meant to take the STRAIN once at baseline, but occasionally some have completed it multiple times and/or may have partial completions, which may result from purposeful or accidental partial completions of the survey (such as if a participant gets 'kicked out' of the survey and has to start over).

This is the dataset most users will want to use as, typically, we would want data from the first administration of the survey to avoid practice effects, etc.

'Partial/Complete Comparisons'

This dataset includes any participant with partial completion data; all participants have at least two entries. This dataset was used to code various flag variables regarding whether these partial completions happened before a full survey completion, etc.

To create an updated version of this file, you will need new partial completion data from Dr. Slavich's team and participant notes from the EAS team.

2.2.5 Take Home Packet

	Snapshot
Project/Core	Clinical Core
Funding period / Vintages	2017–2022; 2023–2028

Filenames in Release

The following file(s) are included in release:

THP_acq_leisure_final_{ts_fn}.csv
THP_discrimination_final_{ts_fn}.csv
THP_emotional_health_final_{ts_fn}.csv
THP_fatigue_final_{ts_fn}.csv
THP_lifeSatisfaction_final_{ts_fn}.csv
THP_loneliness_final_{ts_fn}.csv
THP_neighborhood_final_{ts_fn}.csv
THP_panas_final_{ts_fn}.csv
THP_personality_final_{ts_fn}.csv
THP_physicalfunctioning_final_{ts_fn}.csv
THP_promisSatisfaction_final_{ts_fn}.csv
THP_reap_final_{ts_fn}.csv
THP_socialactivity_final_{ts_fn}.csv
THP_subjectivestress_final_{ts_fn}.csv
THP_support_final_{ts_fn}.csv
THP_wide_{ts_fn}.csv (all packets)

Overview

Take-home questionnaires are self-administered surveys completed by participants outside of the study visit. These instruments assess psychosocial factors such as emotional health, perceived discrimination, stress, and related experiences. These measures provide additional information on participants' social and psychological environments that may influence health and cognitive outcomes.

The Take-Home Questionnaire codebook is now included as a major self-report component, and includes the following measures:

Life Satisfaction	Social Support
Discrimination	Loneliness
Personality	PROMIS Satisfaction
Emotional Health	Leisure / Social Activity
PANAS	Physical Functioning
Subjective Stress	Fatigue
Eating Assessment	Residential History
Neighborhood Quality / Safety	

Note: Variable-level tables are intentionally excluded from this narrative document and are managed in [Airtable](#).

2.3 Sensors

The EAS captures participant data with various wearable sensors that provide minute by minute updates on, for example, a participant's movement, the air quality they experience, or cardiovascular health. The following section includes notes on each sensor included in the study, how they were used, and, where possible, links out to the sensor website for more detailed technical information. Incorporating these sensors into the EAS allows researchers to access granular, measured data on the minute by minute level.

2.3.1 Physical Activity and Sedentary Behavior (Sensor: ActivPAL)

Snapshot	
Project/Core	Project 1
Funding period / Vintages	2023–2028

Filenames in Release

The following file(s) are included in release:

```
QC_activpal_{ts_fn}.csv
```

Overview

The activPAL is a small wearable device that measures body movement and posture. It can distinguish between sitting/lying, standing, and stepping, allowing researchers to quantify patterns of physical activity and sedentary behavior. These data are commonly used to examine mobility, daily activity patterns, and relationships between physical activity and health outcomes.

Physical activity and sedentary behavior are objectively assessed using the activPAL4 micro monitor (PAL Technologies Ltd., Glasgow, UK). This device is lightweight (15g) and houses a 3-dimensional accelerometer that samples movement and posture at 20 Hz. The device provides no direct feedback to users. Consistent with best practices outlined by Edwardson et al. (2017), the monitor will be waterproofed with a nitrile sleeve and attached to the participant’s thigh with a hypoallergenic Hypafix fabric bandage. The monitor can be worn in the shower but participants will be instructed to remove it before taking a bath or swimming. Participants will be given replacement bandages to re-attach the monitor if they opt to remove it temporarily or change the leg to which it is attached.

The monitor has a 16MB capacity and can store at least 14 days of free-living activity data. It classifies time spent in a variety of events/states: non-wear, time-in-bed, sitting [non-transport-related], sitting [transport-related], standing, moving, and cycling (Granat, 2012). For all moving events, the device records step counts and step cadence. Data will be output in event, daily, and person-level files to summarize key features using CREA v.1.3, including duration in each state, the frequency of sit-to-stand transitions, and intensity of movements (≥ 100 steps/min will indicate moderate-to-vigorous intensity physical activity; Tudor-Locke et al., 2018). The activPAL4 is considered the gold-standard measure for sedentary behavior and is comparable if not superior to the waist-worn Actigraph for physical activity (the activPAL is superior to the Actigraph for detecting slow step cadences [Ryan et al., 2006]). Scores for sedentary time and step counts have demonstrated superior accuracy compared to pedometers and waist-worn Actigraph accelerometers (Kozey-Keadle et al., 2011). Rosenberg et al. (2020) found that activPALs were as acceptable to older adults as Actigraph monitors.

ActivPAL Daily-level Export Settings

Related settings notes:

- Wear Time protocol:
 - 24-hour protocol (allow 4 hours non-wear). “valid_day variable =1 when valid (≤ 4 hours non-wear) and = 0 when invalid (> 4 hours non-wear). The minimum duration of any non-wear period is 60 minutes.
- Reciprocal Leg Movements (RLM):
 - Walking, stair climbing, running, and cycling all involve reciprocal leg movements, which can be detected using the thigh sensor location. We use the encompassing term Reciprocal Leg Movements (RLMs) to provide clarity to when the analysis outputs refer to all RLMs, and when they refer to a sub-classification e.g. “walking steps” or “cycling steps”.

software_tool_version	9.1.0.77
validation_algorithm_name	MORA
validation_algorithm_version	v1.0
validation_algorithm_wear_time_protocol	24
analysis_algorithm_name	CREA
analysis_algorithm_version	v1.3
analysis_algorithm_autocorrect_inverted	yes
analysis_algorithm_minimum_upright_seconds	10
analysis_algorithm_minimum_non_upright_seconds	10

CREA Classification Algorithm

Non-wear

- The non-wear detection algorithm is based on a measure of stillness. There are two conditions where the accelerometer signal does not vary for long periods of time, non-wear or when the wearer does not move their leg. The accelerometer is very sensitive to small leg movements but none-the-less it is not uncommon for mobility impaired individuals to be still for up to an hour during waking sitting or during sleep. This version of the non-wear algorithm determines non-wear firstly by identifying the longest blocks of non-varying accelerometer data and then tests adjacent blocks for similar characteristics to build containers of non-wear “activity”.
- Settings (not user adjustable):
 - the minimum duration of non-wear is 60 minutes
- The non-wear containers are used in the validation algorithm.

Upright correction

- For some seated postures, for example perching on a stool or leg positions during lying (e.g. in bed), the intermediate leg angles can cause fluctuations in posture as determined by the VANE algorithm between upright and sitting resulting in false upright detection during non-upright activities. This algorithm corrects for these conditions on a posture container by container basis.

Lying

- This algorithm examines all non-upright events in the calendar day to identify the primary lying period each day.
- For each day, all non-upright events longer than an hour are identified. Each event is then expanded out to adjacent non-upright events (allowing for bathroom breaks / interruptions) resulting in a container of predominantly non-upright events. These containers are then sorted by duration and the longest container flagged as the primary lying container. In most cases this primary container will contain rolling of thigh. In the case the primary container is rolling then the other containers identified and containing rolling will be classed as secondary lying containers. If there is no rolling in the primary lying container then no secondary lying containers can be identified.
- Settings (not user adjustable):
 - the minimum event duration for consideration as lying (primary or secondary) is 60 minutes
 - accumulated upright time in the lying container of >15 minutes will end the container
 - a subsequent rolling event will reset the accumulated upright time counter
 - a sitting bout >15 minutes will end the container
 - where rolling is present in the container, the first and last non-upright events in the container must include rolling
- When the primary lying period contains rolling of the thigh any additional sections of non-upright with rolling are marked as secondary lying.

- Notes:
 - When analyzing the data, the use of secondary lying is context dependent. For example, secondary lying in the middle of the day may reflect someone lying down for a nap. In the evening secondary lying may reflect couch lying (the lying visualization feature in PALanalysis can give a qualitative view of how still wearer is during lying). Depending on the study, these sections may be of specific interest.
 - Another common case is where someone has a long lying period, then gets up in the night (for example not being able to sleep, or attending to family), then goes back to bed for another long lying period which will be most often classified as secondary lying. Secondary lying periods can be joined to the primary lying container using the user-defined feature of the Time in Bed visualization in PALanalysis.
 - For backwards compatibility, secondary lying is included in the sitting totals. Primary lying is not included in “sitting” outputs.

Cycling

- The measurement of thigh inclination provides a robust method to separate cycling leg movements from stepping leg movements. We class all repeated “cyclical” leg movements in an upright posture as Progressive Leg Movements (PLM). The flexed inclination of the thigh during cycling is used as the primary criteria to separate cycling PLM from stepping PLM. The VANE algorithm classifies cycling PLM as stepping, so for backwards compatibility cycling PLM is included in the total step count.
- Settings (not user adjustable):
 - the minimum duration of a cycling bout is 60 seconds
 - more than 50% of the PLM events must meet the criteria for cycling

Seated Transport

- It has been observed that wearable activity monitors may incorrectly classify periods of motorized transport as light or moderate intensity physical activity due to external accelerations generated by, for example, the vehicle’s engine and interactions with the road surface. That is, dynamic accelerations not associated with human movement can result in misclassification of activity type. By using the inclination of the thigh to detect a seated posture we can correctly identify seated car travel as a sedentary behavior regardless of any external accelerations present. Taking this approach one step further we can use the presence of dynamic components in the acceleration signal from a seated subject to identify periods of motorized transport.
- Settings (not user adjustable):
 - minimum transport durations 5 minutes
 - only non-upright events (sitting) can be classed as transport
 - accelerations are assessed in 15 second epochs and where the median noise value across the whole event falls in a moderate noise range the sitting event is classed as transport
 - events are excluded from transport where there are excessive changes in thigh inclination

EAS Data User Notes

- Recommend examining values in variables of interest that are ± 3 SD
- Recommend analyzing data from participants that have at least 4 valid days of data; four days is usually co-investigator Dr. David Conroy’s cutoff for getting a reliable person-level mean
- Important notes on sedentary variables
 - “Primary lying” is defined as the longest non-upright event (longer than an hour) in the day.
 - “Secondary lying” is defined as the sum of any other (not longest) non-upright events (longer than an hour) in the day.
 - Variable `total_sedentary_time_m` is the sum of sedentary categories `sitting_time_m`, `seated_transport_time_m`, and `secondary_lying_time_m`. It does NOT include `primary_lying_time_m`.

- Similarly, the number of and minutes in sedentary bouts (by bout duration) includes the following sedentary categories: sitting time, seated transport time, and secondary lying time. It does NOT include primary lying time.

2.3.2 Air Quality (Sensor: Atmotube Pro)

	Snapshot
Project/Core	Project 3
Funding period / Vintages	2023–2028

Filenames in Release

The following file(s) are included in release:

```
ATMOTUBE_participant_record_counts_{ts_fn}.csv
ATMOTUBE_participant_wave_record_counts_{ts_fn}.csv
ATMOTUBE_pp_w_ids_{ts_fn}.csv
ATMOTUBE_records_per_day_by_pid_wave_{ts_fn}.csv
ATMOTUBE_records_per_day_overall_{ts_fn}.csv
```

Overview

The Atmotube Pro is a portable air quality monitor worn by participants to measure personal environmental exposure. It records levels of airborne pollutants (e.g., volatile organic compounds and particulate matter), as well as temperature and humidity. These measurements allow researchers to study how real-world environmental exposures relate to health, cognition, and daily behavior.

This dataset contains data collected using the Atmotube Pro, a portable air quality monitor. The device measures multiple environmental parameters in real time.

The Atmotube Pro captures the following environmental indicators:

- Volatile Organic Compounds (VOC), measured in parts per million (ppm)
- Temperature, measured in degrees Celsius (°C)
- Relative Humidity, measured as a percentage (%)
- Atmospheric Pressure, measured in millibars (mbar)
- Particulate Matter concentrations:
 - PM1 ($\mu\text{g}/\text{m}^3$)
 - PM2.5 ($\mu\text{g}/\text{m}^3$)
 - PM10 ($\mu\text{g}/\text{m}^3$)

The device may also record geolocation coordinates (latitude and longitude) when available.

For device setup procedures and operational details, refer to the [Atmotube Pro Setup Manual](#).

Data Structure Notes

- Timestamps are recorded in GMT (ISO 8601 format).
- Separate date variables may also be included in MM/DD/YY format.
- Start and end dates indicate the window of data collection for each participant.
- Environmental measurements are recorded as floating-point values.

Time Zone Conversion (GMT to EST)

R Code

```
library(lubridate)

# Example GMT timestamp
gmt_time <- ymd_hms("2024-06-25T17:19:00Z", tz = "GMT")

# Convert to EST (accounts for daylight saving time automatically)
est_time <- with_tz(gmt_time, tzone = "America/New_York")

print(est_time)
```

Python Code

```
from datetime import timedelta

# Fixed EST offset example (does NOT auto-adjust for DST)
est_fixed = gmt_time - timedelta(hours=5)

print(est_fixed)
```

2.3.3 Glucose Monitoring (Sensor: CGM)

	Snapshot
Project/Core	Project 4
Funding period / Vintages	2023–2028

Filenames in Release

The following file(s) are included in release:

```
GLUCOSE_cgm_ids_{ts_fn}.csv
GLUCOSE_summary_by_participant_{ts_fn}.csv
GLUCOSE_summary_by_participant_date_15min_{ts_fn}.csv
GLUCOSE_summary_by_participant_date_{ts_fn}.csv
GLUCOSE_summary_by_participant_date_hour_{ts_fn}.csv
```

Overview

The Abbott Libre 3 is a wearable continuous glucose monitoring (CGM) system that passively measures interstitial glucose levels every 5 minutes throughout the day and night. It provides high-frequency glucose data that enables researchers to examine real-time glycemic patterns and variability in relation to behavior, cognition, and emotional processes in daily life.

This dataset contains glucose data collected using the Abbott Libre 3 CGM system. The Libre 3 sensor is worn on the back of the upper arm and continuously measures glucose levels in interstitial fluid. Measurements are automatically recorded approximately every 5 minutes and transmitted to a paired mobile device.

Summary

The dataset includes time-stamped glucose values along with device and sensor metadata. These data can be used to derive clinically and behaviorally relevant glycemic metrics, including:

- Mean glucose (mg/dL)
- Time in range (TIR; typically 70–180 mg/dL)
- Time above range (TAR) and time below range (TBR)
- Glycemic variability (e.g., SD, CV)

Data are recorded based on the participant’s local device time and are not automatically standardized to a universal timezone. As such, timestamps may shift due to travel across time zones or daylight saving time transitions.

The Libre 3 system operates in real time and does not require manual scanning (unlike earlier Libre versions). However, data completeness may still be affected by device connectivity, sensor wear duration, or technical interruptions.

Notes

- **Timestamp considerations:**
All timestamps reflect the participant’s local phone time. Time changes are automatically updated as phones update the time (e.g., daylight saving transitions), which may result in discontinuities (e.g., skipping from ~1:59 AM to ~3:00 AM).

- **Missing or artifact values:**
Certain values (e.g., repeated codes such as “6” in some exports) may reflect technical artifacts rather than meaningful data. These may arise from app restrictions or device communication issues.
- **Measurement context:**
CGM values reflect interstitial glucose, which lags behind blood glucose by approximately 5–15 minutes. This may want to be considered when aligning CGM data with EMA, cognitive testing, or behavioral events, though we haven’t done this for past publications.
- **Unit of measurement:**
Glucose values are reported in mg/dL.
- **Sensor duration:**
Each Libre 3 sensor is typically worn for up to 14 days. The newest sensor can be worn up to 15 days.

Range checks

- limit of CGM device is 400 mg/dL
- will not log anything above 400 or below 40
- for low end, limit is 40 mg/dL

Common Time Aggregations for Summary Metrics

- Summary metrics (mean glucose, CV, TIR, TIH, TIVH, TIH+TIVH, TIL+TIVL) over 2-week burst (`summary_by_participant_date`)
- Same summary metrics over the 60 minutes prior to cognitive tests or EMA assessment (`summary_by_participant_date_hour`)
- Same summary metrics for 3-hour buckets prior to cognitive tests or EMA assessment (to cover all [or majority] of time between tests; `summary_by_participant_date_3hr`)
- Daytime summary metrics (often based on self-reported or tracked wake time)
- Nighttime summary metrics (often based on self-reported or tracked sleep time)

Common Use of Raw Glucose Values

- Single `Historic_Glucose_mg_d_1` value within 5-15 min before cognitive test start (`summary_by_participant_date_15min`)

2.3.4 Sleep Actigraphy and Oximetry (Sensor: Nonin)

	Snapshot
Project/Core	Project 1; R01
Funding period / Vintages	2017–2022; 2023–2028

Filenames in Release

The following file(s) are included in release:

SENSOR_nonin_ts_fn.csv

Overview

The Nonin pulse oximeter measures blood oxygen saturation (SpO_2) and heart rate. The device typically clips onto a participant's finger and uses light-based sensors to estimate the proportion of oxygenated hemoglobin in the blood. These measurements provide information about cardiovascular and respiratory function and can be used to monitor physiological responses during daily activities.

The Einstein Aging Study Sleep Actigraphy and Nonin (Oximetry) Datasets contain participant-level data that correspond to daily mean sleep actigraphy measures across approximately 2 weeks of data. Individuals who were actively participating in the data collection and did not use continuous positive airway pressure (CPAP) equipment were asked to wear an accelerometer on their non-dominant wrist for 2 weeks (in alignment with the period when individuals were asked to complete EMA sessions) to track their sleep, and an oximeter on the finger that fit best into the sensor (non-dominant hand) for 1 night.

The participants were instructed to wear the watch all the time, day or night, except when the watch could be damaged (participating in contact sports or exposed to extreme temperatures). The watch is water resistant, and participants were told it's fine to wear the watch while shower/bathe, but to dry the watch and skin under the watch when finished.

For individuals who participated in the data collection with ambulatory device, not all of the collected data passed the data quality check. For detailed information, please refer to Actigraphy valid days, and Nonin valid recordings, for details.

Valid data and merge with other data sources

The day-level sleep actigraphy data set includes activity and light level measurements for all days that the sleep watch was collecting data, even if the participant was not wearing it (refer to the “valid days” section in the documentation). The day-level flag variable “valid_day” indicates whether a given 24-hour day is considered valid. For analyses, it is advisable to use sleep actigraphy data from participants with at least 4 valid days. A greater number of valid days for an individual yields more accurate estimates of that individual's regular sleep patterns. However, each study should consider appropriate sensitivity analyses to justify any specific cut-off criteria.

Note that as some of the participants wore the sleep actigraphy device beyond the duration specified in the study protocol, it is common and highly recommended to discuss with the study team to decide how many days of the sleep actigraphy data should be included in the analyses. If the data user wishes to align the sleep actigraphy data with other data sources from the same time period, it is recommended to filter appropriate valid days by date from the day-level data set to ensure consistency with the corresponding data set.

We recommend using sleep actigraphy data and Nonin data together when conducting any data analysis, as sleep disordered breathing (SDB, detailed variables listed in Nonin data dictionary) is usually considered as a covariate. It is also highly recommended to check the general EAS documentation regarding other inclusion/exclusion criteria (e.g., dementia) when using the sleep and Nonin datasets.

Number of participants included in data sets by burst

The day-level sleep actigraphy and mean-level Nonin data sets each include all participant bursts combined into a single data file. To filter on a specific participant or a specific burst, please use variables `participant_id` and `burst`.

R01 Number of participants included in data sets by burst		
	Sleep	Nonin
Burst 1	311	296
Burst 2	204	196
Burst 3	193	187
Burst 4	159	154
Burst 5	89	89
Burst 6	28	28

P01 Number of participants included in data sets by burst		
	Sleep	Nonin
Burst 1	257	250
Burst 2	132	132
Burst 3	41	45

File Layout

The Sleep Actigraphy and oximeter data are released in two datasets with different formats.

Day-level Sleep Actigraphy Dataset

- One row per day per participant per burst; this file is sorted by the identifiers `participant_id` and `burst`.
- Detailed data dictionary available in [Airtable](#).

Nonin Oximetry Dataset

- One row per participant per burst; this file is sorted by the identifiers `participant_id` and `burst`.
- Detailed data dictionary available in [Airtable](#).

Variable Naming Convention

This section provides an overview of how variables are named in the above-mentioned datasets, and what are the major categories each variable fall into.

The Sleep Actigraphy variables in these datasets follow the same set of naming rules. In summary, the datasets provided Sleep Actigraphy information regarding 24-hour sleep, nighttime sleep, and daytime naps.

The oximeter variables provided in the datasets provide information regarding hypoxemia classification, sleep disordered breathing (`texttttsdb_odi`) classification, and frequency of desaturation.

Actigraphy variable naming convention

Actigraphy variable names are up to 27 characters long. The first 4 characters may contain the variable prefix `dact_` to indicate these are constructed actigraphy variables. The remaining characters indicate the type of variable measure, which include timing and duration variables, 24-hour level measures or nighttime-only measures. Variable definitions are detailed in the variable definitions in the “Detailed description of sleep variable categories” and “Data Dictionary” sections of this document. Below are some of the most common or crucial variable abbreviations.

Oximeter variable naming convention

Oximeter variable names are up to 17 characters long. Variable definitions are detailed in the variable definitions in the “Detailed description of nonin variable categories” and “Data Dictionary” in the [Airtable](#). Below are some of the most common or crucial variable abbreviations.

Character	Indicates
dailysleep	Variable describes all sleep periods within a 24-hour period
nightsleep	Variable describes only nighttime (or the longest) sleep period
nap	Variable describes nap patterns and duration
dec	Decimal time
_c	Decimal time is midnight-centered (e.g., midnight is 0, 1 AM is 1)
mins	Variable unit is in minutes

Character	Indicates
_3p	The variable is calculated based on 3% data
_4p	The variable is calculated based on 4% data
_spo2	Oxygen Saturation
_odi	Oxygen Desaturation Index
_sdb	Sleep Disordered Breathing

Missing Data Codes

Missing data values in the sleep actigraphy data were left blank to avoid confusion. When using this data set, it is recommended to first filter the data on participants who provide at least 4 valid days of sleep actigraphy (for detailed explanation see Valid Days), and exclude individuals who don't have valid nonin data (`analyzed_3hr_valid = 1`).

Important Filtering Variables and Flag Variables

There are 2 filter/flag variables for the sleep data, and 1 filter variable for the oximeter data.

- **Actigraphy Valid Day:** `valid_day`
 - Whether this day of Sleep Actigraphy data is valid.
 - Values: 1 = Yes, 0 = No.
- **Actigraphy “Allnighter”:** `allnighter`
 - Indicates whether the participant had no scored sleep periods during the 24-hour day.
 - Values: 1 = Yes, 0 = No.
- **Nonin Valid Data:** `analyzed_3hr_valid`
 - Nonin valid recording flag indicating whether a participant had at least 3 hours of valid Nonin recording.
 - Values: 0 = Less than 3 hours, 1 = At least 3 hours.
 - It is recommended to only use participant data with valid oximetry recordings (`analyzed_3hr_valid = 1`).

Data Methods

Sleep Actigraphy

Actigraphy device, software, and initial processing

Sleep actigraphy data were collected at 30-second epochs with a wrist-worn accelerometer (Actiwatch Spectrum; Philips-Respironics, Murrysville, PA) worn on a participant's non-dominant wrist, day and night, for 14 days. The devices were given to the participants during their second clinical visit at the Einstein Medical Center, and were returned during their third clinical visit. From March 2020 to November 2022, data collection was conducted remotely due to the COVID-19 protocol, with devices distributed and re-collected via mail.

Staff at the Einstein Medical Center downloaded the actigraphy recording from each device using Philips Actiware software version 6.1.2 and shared via Box with staff in the Sleep, Health, and Society Collaboratory (SHSC) at Penn State. Staff in the SHSC exported the 30-second epoch data from Actiware 6.1.2 to CSVs in preparation for scoring.

The medium sensitivity wake threshold option in the software (40 counts per minute) was selected in calculating sleep variables.

Scoring Methods

At least two independent, trained scorers reviewed and visually scored each recording using a standard validated algorithm (see 2013 Marino et al. Sleep; DOI: 10.5665/sleep.3142) in a graphical user interface. Scorers determined cut-point times, validity of days, and set sleep intervals, without using information from a sleep diary.

The cut-point selected for each recording determines the “start” and “end” of a 24-hour day. The preferred cut-point is at noon for each recording; however, the cut-point can be shifted (as close to noon as possible) to select a time that intersects the minimum number of sleep periods and off-wrist periods in a recording. Scorers determined sleep intervals using a decrease in activity levels and the aid of light levels for sleep onset and sleep offset. The main/nighttime sleep period was typically scored between 8pm and 8am and was usually consolidated- the nighttime sleep interval was not split into multiple sleep periods (night sleep and nap) if there was an awakening ≥ 1 hour during this time period. Sleep intervals were not scored if the duration of an interval was less than 20 minutes; therefore, any nap or nighttime sleep duration must be greater or equal to 20 minutes.

After individual scoring was completed, the scorers adjudicated each recording for interrater agreement by verifying number of valid days, cut-point, number of sleep intervals (night sleep and naps), and differences greater than 15 minutes in duration and wake after sleep onset (WASO) for each sleep interval.

Valid Days

The accelerometer had an on-wrist detection feature that allowed scorers to view when participants were not wearing the device. A sleep actigraphy day was determined invalid and no sleep interval was set if there were ≥ 4 total hours of off-wrist time, with the exception of the first and last day (device should be worn at least 2 hours before sleep onset on the first day), constant false activity due to battery failure, or an off-wrist period of ≥ 60 minutes within 10 minutes of the scored beginning or end of the night sleep period for that day. For analyses, it is recommended to use data for participants who have at least 4 valid days. A greater number of valid days for an individual provides better mean estimates of that individual’s regular sleep patterns. However, each study may wish to consider appropriate sensitivity analyses to justify any specific cut-off choices.

Sleep Variable Categories

Overview

Nighttime sleep measures, such as timing, duration, TST (total sleep time), rest WASO (wake after sleep onset), and sleep maintenance efficiency only include data from what is considered the participant’s “nighttime sleep interval”. The nighttime sleep interval duration was calculated as the number of minutes between sleep onset and sleep offset during the sleep interval, which was defined as the sleep interval with the longest duration between the hours of 10PM and 8AM in a 24-hour cut-point day. All other sleep intervals within the 24-hour cut-point day were considered naps and were not included in the nighttime sleep variable measures.

All constructed sleep variables (variables with the `dact_` prefix), with the exception of the flag variables and date variables, can fall into 1 of 9 categories of sleep measures detailed below. Variable categories are indicated by “nighttime” or “24-hr/daily” or both. Exact variable names are listed below each category and exclude the `dact_` prefix in this section.

Detailed description of sleep variable categories

Note: italicized variables are variables unique to the day-level dataset. The `dact_prefix` is not listed below (e.g., for `nightsleep_start_dec_c` listed below, the full variable name in the dataset is `dact_nightsleep_start_dec_c`)

1. Valid days

- In the day-level sleep actigraphy dataset, there is a binary variable that indicates whether a day is valid (=1) or not (=0). Data users should only use valid days in analyses.

Variable: `valid_day`

2. Sleep onset timing (nighttime)

- Sleep onset was defined as the nighttime sleep duration start time: the time of the last 30-second epoch of activity >10 counts followed by 5 consecutive epochs ≤ 10 , indicating the first epoch of sleep.
- The centered sleep onset timing variable was constructed as midnight-centered decimal time. For example, the time “0.00” indicates midnight/12:00AM, “-1.20” indicates 10:48PM (or 1.2 hours before midnight), and “2.45” indicates 2:27AM (or 2.45 hours after midnight). **This centered onset timing variable is typically the appropriate variable to use for sleep onset timing analyses.**

Variable: `nightsleep_start_dec_c`

- Another type of sleep onset timing variable was constructed based on actual decimal time (not midnight-centered).

Variable: `nightsleep_start_dec`

- The date (MM/DD/YYYY), day of week (Sun-Sat), and time (HH:MM:SS) of sleep onset is also included in the day-level dataset.

Day-level data specific variables: `nightsleep_startdate`, `nightsleep_starttime`, `nightsleep_weekday`

3. Sleep offset timing (nighttime)

- Sleep offset was defined as the nighttime sleep duration end time: the time of the first 30-second epoch with activity count >10 preceded by 5 consecutive 30-second epochs ≤ 10 , indicating the last epoch of sleep.

- The centered sleep offset timing variable was constructed as midnight-centered decimal time

Variable: `nightsleep_end_dec_c`

- Another type of sleep offset timing variable was constructed based on actual decimal time (not midnight-centered). **This set of offset timing variables is typically the appropriate set to use for sleep offset timing analyses.**

Variable: `nightsleep_end_dec`

- The date (MM/DD/YYYY), time (HH:MM:SS), day of week in integer form (where 1 = Sunday, 2 = Monday, ..., 7 = Saturday), and whether the day of week was on a weekend (Saturday or Sunday) of sleep offset is also included in the day-level dataset.

Day-level data specific variables: `nightsleep_enddate`, `nightsleep_endtime`, `nightsleepend_weekday`

4. Sleep midpoint timing (nighttime)

- Sleep midpoint was defined as the time halfway between sleep onset and sleep offset during the nighttime sleep duration interval. The sleep midpoint timing variable was constructed as **midnight-centered decimal time**.

Variable: `nightsleep_mid_dec`

5. Sleep duration (nighttime and 24-hr/daily)

- Sleep duration is calculated as the total number of minutes between sleep onset and sleep offset in a sleep interval, including any wake time (minutes of WASO). Nighttime sleep duration (`nightsleepdur`) includes the number of minutes between sleep onset and sleep offset during the nighttime sleep interval only. 24-hour/daily sleep duration (`dailysleepdur`) includes the number of minutes in the nighttime sleep interval (`nightsleepdur`) plus any nap minutes within a 24-hr cut-point day.

Variables: `totalsleepdur_mins`, `nightsleepdur_mins`

6. Total sleep time (nighttime and 24-hr/daily)

- Total sleep time (TST) is calculated as the total number of minutes that are considered sleep between sleep onset and sleep offset in a sleep interval, and does not include any wake time (WASO). Nighttime TST (`nighttst`) includes the number of minutes of sleep between sleep onset and sleep offset during the nighttime sleep interval only. 24-hour/daily TST (`dailytst`) includes the number of minutes of sleep in the nighttime sleep interval (`nighttst`) plus any nap minutes within a 24-hr cut-point day.

Variables: `totalsleeptime_mins`, `nightsleeptime_mins`

7. Wake after sleep onset - WASO (nighttime)

- WASO represents the number of minutes of wake between sleep onset and sleep offset during the nighttime sleep interval. The calculation of this variable is: $\text{restwaso} = \text{nightssleepdur} - \text{nighttst}$. WASO is typically used as a measure of sleep quality; increased WASO indicates lower sleep quality.

Important Note: Please only use the WASO variables that start with “restwaso”.

Variable: `restwaso_mins`

8. Sleep maintenance efficiency (nighttime)

- Sleep maintenance efficiency (`smeff`) was defined as the percentage of minutes (unit: 0-100) of total sleep time (`nighttst`) between sleep onset and sleep offset in the nighttime sleep duration interval (`nightssleepdur`). The calculation of this variable is: $\text{smeff} = (\text{nighttst} / \text{nightssleepdur}) * 100$. Sleep maintenance efficiency is typically used as a measure of sleep quality; higher sleep maintenance efficiency indicates better quality sleep.

Variable: `smeff`

9. Naps (24-hr/daily)

- Nap measures include any sleep intervals in a 24-hr cut-point day that are not the nighttime sleep interval. The nap variables include: the total minutes per day of nap duration (i.e. `nap_mins`), the total number of naps in the day (`nap_n`), and the proportion of nap minutes out of total rest minute (nap and main sleep) in the day (`nap_percent`).

Variables: `nap_mins`, `nap_n`, `nap_percent`, `nap_waso_mins`

Note: Individual nap duration and timing (not summed across the day) are available in variables 73-108.

Nonin Oximetry

Nonin device, software, and initial processing

Nonin oximetry data were collected at 1-second level, with a wrist-worn device and a finger sensor attached to it (Nonin Medical Inc, Plymouth MN). The wrist-worn device is worn on a participant’s non-dominant wrist during the last night when they wear the Actigraphy device, and the finger sensor is worn on the finger on the non-dominant hand that fits best into the sensor. The devices were given to the participants during their second clinical visit at the Einstein Medical Center, and were returned during their third clinical visit. Staff at the Einstein Medical Center downloaded the Nonin recording from each device using nVision 6.5.1.2 and shared via Box with staff in the Sleep, Health, and Society Collaboratory (SHSC) at Penn State. Staff in the SHSC scored the data (link to scoring section) in nVision 6.5.1.2, and exported the scored data from nVision to PDFs in preparation for the final dataset.

For detailed information regarding the device, please refer to nVision and WristOx2 training: [Nonin nvision training](#)

Scoring Methods

At least two independent, IRB-approved, trained scorers are involved in the scoring process. Scoring training involves several meetings to discuss & review example files with an RPSGT (M.M.Gray (Schade) Ph.D., Buxton team) until satisfactory distinction of artifact vs quality data is accomplished in the opinion of the trainer.

Scorers then review and visually score each recording in nVision 6.5.1.2. Scoring involves identifying artifacts in the data and manually excluding those segments from Nonin automated analysis, so that summary statistics in the output reflect only recorded data of sufficient quality. A first Scorer completes this process and saves the new file; a second Scorer performs a random 5%-10% spot-check.

Additionally, some files are identified as “difficult” by the first Scorer, and the second Scorer independently scores each of the “difficult” files identified. To reconcile “difficult” files, the two scorers meet and review together any differences in the fully-scored files (“differences” being substantially different total analyzed data time, or visually different segments of data selected for exclusion). Any remaining uncertainties after adjudication are then reconciled by meeting with the original trainer (Gray) and, if needed (as determined by the trainer), are further reviewed with a consulting clinician on the EAS team (S. Bertisch, MD, MPH).

Valid recordings

The variable `Analyzed` in the dataset displays the total valid recording time (after the manual exclusion of artifactual segments) for each participant. For analysis, it is recommended to use data for participants who have at least 3 hours of analyzed recording time. This criterion is available to filter, using the variable `analyzed_3hr_valid = 1`.

Nonin variable categories

`Variables_3p_spo2_index_1_per_hr` and `_4p_spo2_index_1_per_hr` each reflect an index (/hr) of the number of oxygen desaturation events recorded by Nonin. A drop in SpO₂ of at least 3% (`_3p`) or 4% (`_4p`), each for a minimum duration of 10 seconds, is available depending upon the desired desaturation threshold.

`Variables_3p_odi_sdb` and `_4p_odi_sdb` each reflect whether the participant met a selected threshold criterion of Sleep-disordered Breathing (SDB). In this case, the threshold is 15 events /hr on average, or the clinical threshold used to identify the border between “mild” and “moderate” sleep apnea, if the odi (Oxygen Disturbance Index) were interpreted as an ahi (Apnea Hypopnea Index). That is: based on whether a 3% or 4% oxygen desaturation event criterion was selected, did a given participant meet or exceed a threshold of 15 events per hour? If so, then that participant meets the criterion for SDB+ (coded as = 1)

Variable `time_1t_88pct` represents the total amount of time (in minutes) during data collection where an individual has a recorded peripheral blood oxygenation (SpO₂) less than 88%. This threshold is relevant because it is the criterion used by Medicare to evaluate qualification for supplemental oxygen insurance coverage.

`hypoxemia_88pct_sdb` is a hypoxemia classification variable. Individuals who have `time_1t_88pct > 5` minutes meet the Hypoxemia+ SDB criterion (coded as = 1)

EAS Ambient Light Dataset

This dataset was derived from the epoch-by-epoch light Actiwatch Spectrum Plus data. The Spectrum Plus logs white, red, green, and blue light in 30-second epochs.

The white light channel measures illuminance, and is output in units of lux. The red, green, and blue (RGB) channels measure irradiance, and are output in units of microwatts per square centimeter.

In this dataset, the light values are collapsed from 30-second epochs into day intervals in a combination of ways according to the day cut-point and epoch validity criteria.

Day cut-point

Separate files contain data defined by two different day cut-points.

- 1) The file `Light_Levels_by_Sleep_Day_Export` contains days defined by the same day cut-points used for the actigraphy defined sleep intervals.
- 2) The file `Light_Levels_by_Calendar_Day_Export` contains days defined by midnight-to-midnight intervals.

Epoch Selection for Inclusion

Within each file, several sets of output values are provided with different criteria for the 30-second epochs being selected for inclusion in the day summary:

- 1) All epochs: All available epochs within the day are selected for the light statistics
- 2) Onwrist epochs: All epochs within the day that are marked as ‘on wrist’ (the participant is wearing the Spectrum Plus) are selected.
- 3) Onwrist, no main rest epochs: All epochs within the day that are marked as ‘on wrist’ and that are not within the bounds of a main rest interval (the primary sleep interval identified from the sleep analysis) are selected.
- 4) Onwrist, no main rest or nap epochs: All epochs within the day that are marked as ‘on wrist’ and that are not within the bounds of a main rest interval, or any nap intervals, are selected.

Note that the rest intervals may include some time where the participant is awake. See the sleep actigraphy documentation for more detail on how the rest intervals are defined.

Summary Data Per Epoch:

For each day, the selected epochs are summarized separately for white light, red light, green light and blue light. Each color of light is summarized by:

- 1) `num_samples`: The number of samples selected within the day. This count can sometimes vary slightly by light color, as for example white light can sometimes be NaN while colored light has a valid sample.

- 2) **mean**: The mean of selected light values within the day.
- 3) **sd**: The standard deviation of selected light values within the day (with denominator of n-1).
- 4) **min**: The minimum light value within the day.
- 5) **max**: The maximum light value within the day.
- 6) **sum**: The sum of light values within the day.
- 7) **exposure**: The sum of light values within the day, per minute. As each epoch represents 30-second of data, this is the epoch sum / 2.

Variables:

Each file contains identifying metadata, as well as the light columns assembled from a combination of epoch selection and summary value described above.

In the sleep day defined cutpoint file, the metadata columns include:

- 1) **participant_id**: The EAS participant identifier
- 2) **burst**: The EAS burst identifier
- 3) **intervaln**: The interval number of the day, corresponding to the same variable in the EAS sleep dataset.
- 4) **start_datetime**: The date and timestamp for the beginning of the cutpoint day.
- 5) **end_datetime**: The date and timestamp for the end of the cutpoint day. Note that like the EAS sleep data, this end timestamp not exclusive, the data included for the day includes everything up to, but not including, the end datetime.

In the midnight-to-midnight defined cutpoint file, the metadata columns include:

- 1) **participant_id**: The EAS participant identifier
- 2) **burst**: The EAS burst identifier
- 3) **date**: The date for which the data belongs. Samples summarized for the day are from 12:00 AM to 11:59:30 PM on this date.

For both files, the remaining columns represent a combination of epoch selection and summary value. For example, the column `onwrist_epochs_white_light_mean` represents using the `onwrist_epochs`, is summarizing the `white_light`, and is the mean value across the day.

Part 3. References

Blood Biomarkers

Inflammation

Van Bogart, K., Engeland, C. G., Sliwinski, M.J., Harrington, K.D., Knight, E.L., Zhaoyang, R., Scott, S.B., & Graham-Engeland, J.E. (2022). The association between loneliness and inflammation: Findings from an older adult sample. *Frontiers in Behavioral Neuroscience*, 15: 801746. doi: 10.3389/fnbeh.2021.801746

Knight, E. L., Majd, M., Graham-Engeland, J. E., Smyth, J. M., Sliwinski, M. J., & Engeland, C. G. (2020). Gender differences in the link between depressive symptoms and ex vivo inflammatory responses are associated with markers of endotoxemia. *Brain, Behavior, & Immunity-Health*, 2, 100013.

Fagundes, C. P., Brown, R. L., Chen, M. A., Murdock, K. W., Saucedo, L., LeRoy, A., ... & Heijnen, C. (2019). Grief, depressive symptoms, and inflammation in the spousally bereaved. *Psychoneuroendocrinology*, 100, 190-197.

SNP Genotyping - APOE Genotypes

Pihlström et al *Neurobiol Aging*. 2013;34:1708.e7-13.

Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease. Shulman et al *JAMA Neurol*. 2014 Feb 10.

Association of Parkinson Disease Risk Loci With Mild Parkinsonian Signs in Older Persons.

(Citation for Imputation Server): Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze S, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. Next-generation genotype imputation service and methods. *Nature Genetics* 48, 1284–1287 (2016.)

Clinical Core

STRAIN

Barzilai, N., Rossetti, L., & Lipton, R. B. (2004). Einstein's Institute for Aging Research: collaborative and programmatic approaches in the search for successful aging. *Experimental Gerontology*, 39(2), 151–157. <https://doi.org/10.1016/j.exger.2003.10.009>

Crystal, H. A., Dickson, D., Davies, P., Masur, D., Grober, E., & Lipton, R. B. (2000). The Relative Frequency of Dementia of Unknown Etiology Increases With Age and Is Nearly 50% in Nonagenarians. *Archives of Neurology*, 57(5), 713–719. <https://doi.org/10.1001/archneur.57.5.713>

Katz, M. J., Lipton, R. B., Hall, C. B., Zimmerman, M. E., Sanders, A. E., Verghese, J., Dickson, D. W., & Derby, C. A. (2012). Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and Alzheimer dementia in blacks and whites: a report from the Einstein Aging Study. *Alzheimer Disease and Associated Disorders*, 26(4), 335–343. <https://doi.org/10.1097/WAD.0b013e31823dbcf>

Sensors

activPAL

Edwardson CL, Winkler EAH, Bodicoat DH, Yates T, Davies MJ, Dunstan DW, Healy GN. Considerations when using the activPAL monitor in field-based research with adult populations. *J Sport Health Sci*. 2017 Jun;6(2):162-178. doi: 10.1016/j.jshs.2016.02.002. Epub 2016 Feb 3. PMID: 30356601; PMCID: PMC6188993.

Granat MH. Event-based analysis of free-living behaviour. *Physiol Meas*. 2012 Nov;33(11):1785-800. doi: 10.1088/0967-3334/33/11/1785. Epub 2012 Oct 31. PMID: 23110873.

Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc*. 2011 Aug;43(8):1561-7. doi: 10.1249/MSS.0b013e31820ce174. PMID: 21233777.

Rosenberg D, Walker R, Greenwood-Hickman MA, Bellettiere J, Xiang Y, Richmire K, Higgins M, Wing D, Larson EB, Crane PK, LaCroix AZ. Device-assessed physical activity and sedentary behavior in a community-based cohort of older

adults. *BMC Public Health*. 2020 Aug 18;20(1):1256. doi: 10.1186/s12889-020-09330-z. PMID: 32811454; PMCID: PMC7436994.

Ryan CG, Grant PM, Tigbe WW, Granat MH. The validity and reliability of a novel activity monitor as a measure of walking. *Br J Sports Med*. 2006 Sep;40(9):779-84. doi: 10.1136/bjsm.2006.027276. Epub 2006 Jul 6. PMID: 16825270; PMCID: PMC2564393.

Tudor-Locke C, Han H, Aguiar EJ, Barreira TV, Schuna JM Jr, Kang M, Rowe DA. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. *Br J Sports Med*. 2018 Jun;52(12):776-788. doi: 10.1136/bjsports-2017-097628. PMID: 29858465; PMCID: PMC6029645.

Glucose Monitoring

Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, Bosi E, Buckingham BA, Cefalu WT, Close KL, Cobelli C. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes care*. 2019;42(8):1593-603.

Danne T, Nimri R, Battelino T, Bergenstal RM, Close KL, DeVries JH, Garg S, Heinemann L, Hirsch I, Amiel SA, Beck R. International consensus on use of continuous glucose monitoring. *Diabetes care*. 2017;40(12):1631-40.

Beck RW, Bergenstal RM, Riddlesworth TD, Kollman C, Li Z, Brown AS, Close KL. Validation of time in range as an outcome measure for diabetes clinical trials. *Diabetes care*. 2019 Mar 1;42(3):400-5.

Sleep/Nonin

Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2012;8(5):597-619. Published 2012 Oct 15. doi: 10.5664/jcsm.2172

Vaughan L, Redline S, Stone K, et al. Feasibility of self-administered sleep assessment in older women in the Women's Health Initiative (WHI). *Sleep Breath*. 2016;20(3):1079-1091. doi:10.1007/s11325-016-1314-3

Part 4. Appendix

4.1 SNPs for Genotyping

Return to Section [2.1.4](#)

ITEM	GENE	SNP ACCESSION #	OTHER CONVENTIONS	RATIONALE
TOP 10 list from ALZgene	APOE e2/3/4	several; rs429358*; C=0.149; Rs7412*; T=0.073		AD risk and susceptibility; too many references to list. *Previously genotyped in EAS
TOP 10 list from ALZgene	B1N1	rs744373; CH2q14; Global MAF; G=0.3714; African American sp.; Rs55636820; MAF A=0.03; Intron variant	also known as AMPH2;AMPHL	Bridging integrator 1- encodes several isoforms of nucleocytoplasmic adaptor protein. Isoforms that are expressed in the central nervous system may be involved in synaptic vesicle endocytosis and may interact with dynamin, synaptojanin, endophilin, and clathrin. Alternate splicing of the gene results in ten transcript variants encoding different isoforms.
TOP 10 list from ALZgene	CLU	rs11136000; CH8; Global MAF; T=0.3848; Intron variant	CLI; APOJ; TRPM-2; TRPM2; SP-40; SGP-2; SGP2; MGC24903	Clusterin - secreted chaperone that can be found in the cell cytosol, involved in cell death, tumor progression, and neurodegenerative disorders
TOP 10 list from ALZgene	ABCA7	rs3764650; CH19; Global MAF; G=0.1997; Intron variant; African American sp.; Rs115550680; G=0.018; Intron variant	ABCA-SSN; ABCX; FLJ40025	ATP-binding cassette sub-family A member 7 - member of the superfamily of ATP-binding cassette (ABC) transporters; transport various molecules across extra- and intra-cellular membranes. Function not clear, expression pattern suggests a role in lipid homeostasis in cells of the immune system. (Baranzini 2008, Haraquni 2005)
TOP 10 list from ALZgene	CR1	rs3818361; CH1; Global MAF; A=0.2691; Intron variant; African American sp.; Rs146366639; MAF 0.26	C3BR; C4BR; CD35; KN	Complement component c3/4b receptor 1 - \member of the receptors of complement activation (RCA) family; mediates cellular binding to particles and immune complexes that have activated complement.
TOP 10 list from ALZgene	PICALM	rs3851179; CH11; Global MAF; T=0.3297	CALM; CLTH; LAP	Phosphatidylinositol binding clathrin assembly protein - encodes a clathrin assembly protein, which recruits clathrin and adaptor protein complex 2 (AP2) to cell membranes. The protein may be required to determine the amount of membrane to be recycled, possibly by regulating the size of the clathrin cage.
TOP 10 list from ALZgene	MS4A6A	rs610932; CH11; Global MAF; T=0.4219; utr variant 3'	CDA01; 4SPAN3; 4SPAN3.2; CD20L3; MGC131944; MGC22650; MS4A6; MST090; MSTP09	membrane-spanning 4-domains, subfamily A, member 6A - gene encodes a member of the membrane-spanning 4A gene family. Members of protein family are characterized by common structural features and similar intron/exon splice boundaries and display unique expression patterns in hematopoietic cells and nonlymphoid tissues.
TOP 10 list from ALZgene	CD33	rs3865444; CH19; Global MAF; A=0.2378; upstream variant 2KB; African American sp.; Rs114282264; MAF 0.03; Intron variant	p67; SIGLEC3; FLJ00391; SIGLEC-3	CD33 - is a sialic acid-binding immunoglobulin-like lectin that regulates innate immunity; inhibits microglial uptake of amyloid beta
TOP 10 list from ALZgene	MS4A4E	rs670139; Global MAF; T=0.3999	CH11	membrane-spanning 4-domains, subfamily A, member 4E - encodes proteins with at least 4 potential transmembrane domains and N- and C-terminal cytoplasmic domains encoded by distinct exons
TOP 10 list from ALZgene	CD2AP	rs9349407; CH6; Global MAF; C=0.1905; Intron variant	CMS; DKFZp586H051	CD2-associated protein - encodes a scaffolding molecule that regulates the actin cytoskeleton; implicated in dynamic actin remodeling and membrane trafficking that occurs during receptor endocytosis and cytokinesis.

ITEM	GENE	SNP ACCESSION #	OTHER CONVENTIONS	RATIONALE
GWAS neuritic plaque studies	KCNIP4 – Of secondary priority	rs6817475; Global MAF; G=0.3076; Intron variant	intronic	encoding a potassium channel-interacting protein. Notably, KCNIP4 physically interacts with PSEN235 and alters A β dynamics in cultured cells ³² ; further, insertion deletion polymorphisms in the KCNIP4 promoter were associated with AD in a small case-control autopsy cohort
GWAS neuritic plaque studies	PTGS1	rs12551233; Global MAF; G=0.1928; Intron variant		also known as cyclooxygenase 1 (COX1), which encodes a key regulator of inflammation. Increased COX1 along with other inflammatory markers have been described in association with neuritic plaque pathology
GWAS neuritic plaque studies	ATP5J-APP - Of secondary priority	rs2829887; Global MAF; T=0.3370; Intron variant	intronic	Proxymal to APP
GWAS neuritic plaque studies	NMNAT3 - Of secondary priority	rs4564921; Global MAF; C=0.4086		nicotinamide nucleotide adenylyltransferase 3 gene This enzyme family has demonstrated neuroprotective activities in experimental models, ³⁷ and the NMNAT3 locus was previously implicated in an AD genome-wide scan
GWAS neuritic plaque studies	SLC35F4 - Of secondary priority	rs187911; Global MAF; G=0.3765		solute carrier family 35, member F4 – associated with bipolar disorder
GWAS neuritic plaque studies	NPAS3	rs10149826; Global MAF; T=0.0882		encodes a member of the basic helix-loop-helix and PAS domain-containing family of transcription factors. The encoded protein is localized to the nucleus and may regulate genes involved in neurogenesis.
GWAS neuritic plaque studies	PARD3B - Of secondary priority	rs12613305; Global MAF; A=0.3577		par-3 family cell polarity regulator beta is a protein-coding gene. associated lateral sclerosis, and ALS
From IGAP Previously	BIN 1 (old)	rs6733839; Global MAF; T=0.4082	C/T	box-dependent-interacting protein 1 - expressed in the central nervous system may be involved in synaptic vesicle endocytosis and may interact with dynamin, synaptoanin, endophilin, and clathrin
From IGAP Previously	EPHA1 (old)	rs11771145; Global MAF; A=0.4596; Intron variant; African American sp.; Rs6973770; MAF G = 0.06; From Reitz et al.	G/A	Erythropoietin-Producing Hepatoma - ephrin receptor subfamily of the protein-tyrosine kinase family. EPH and EPH-related receptors have been implicated in mediating developmental events, particularly in the nervous system
From IGAP Previously	HLA-DRB5–HLA-DRB1	rs9271192; Global MAF; C=0.2401; Rs3129882 - Laurie	A/C	encoding major histocompatibility complex, class II, DR β 5 and DR β 1, respectively. This region is associated with immunocompetence and histocompatibility and with risk of both multiple sclerosis and Parkinson disease
From IGAP Previously	PTK2B	rs28834970; Global MAF; C=0.2792; Intron variant	T/C	protein tyrosine kinase 2 β - involved in the induction of long-term potentiation in the hippocampal CA1 (cornu ammonis 1) region, a central process in the formation of memory
From IGAP Previously	SLC24A4-RIN3	rs10498633; Global MAF; T=0.1543; Intron variant	G/T	solute carrier family 24 (sodium/potassium/calcium exchanger), member 4 - involved in iris development and hair and skin color variation in humans in addition to being associated with the risk of developing hypertension; also expressed in the brain and may be involved in neural development
IGAP New loci reaching gwa significance in discovery and replication analyses	INPP5D	rs35349669; Global MAF; T=0.2475	C/T	inositol polyphosphate-5-phosphatase - expressed at low levels in the brain, but the interacts with CD2AP, whose corresponding gene is one of the Alzheimer's disease genes and modulates, along with GRB2, metabolism of APP

ITEM	GENE	SNP ACCESSION #	OTHER CONVENTIONS	RATIONALE
IGAP New loci reaching gwa significance in discovery and replication analyses	MEF2C	rs190982; Global MAF; G=0.2185	A/G	myocyte enhancer factor - Mutations at this locus are associated with severe mental retardation, stereotypic movements, epilepsy and cerebral malformation; limits excessive synapse formation during activity-dependent refinement of synaptic connectivity and thus may facilitate hippocampal-dependent learning and memory
IGAP New loci reaching gwa significance in discovery and replication analyses	NME8	rs2718058; Global MAF; G=0.3586	A/G	encoding NME/NM23 family member - responsible for primary ciliary dyskinesia type 6
IGAP New loci reaching gwa significance in discovery and replication analyses	ZCWPW1	rs1476679; Global MAF; C=0.2323; Intron variant	T/C - INTRONIC	encoding zinc finger, CW type with PWWP domain 1) - corresponding protein modulates epigenetic regulation
IGAP New loci reaching gwa significance in discovery and replication analyses	CELF1	rs10838725; Global MAF; C=0.2218; Intron variant	T/C	encoding CUGBP, Elav-like family member 1, member of the protein family that regulates pre-mRNA alternative splicing
IGAP New loci reaching gwa significance in discovery and replication analyses	FERMT2	rs17125944; Global MAF; C=0.1226; Intron variant	T/C	fermitin family member 2 -expressed in the brain. localizes to cell matrix adhesion structures, activates integrins, is involved in the orchestration of actin assembly and cell shape modulation, and is an important mediator of angiogenesis
IGAP New loci reaching gwa significance in discovery and replication analyses	CASS4	rs7274581; Global MAF; C=0.1088 ; Intron variant	T/C	Cas scaffolding protein family member 4 - Little is known about the function, but the DrosophilaCASS family ortholog (p130CAS) binds to CMS, the Drosophila ortholog of CD2AP is a known Alzheimer's disease susceptibility gene that is involved in actin dynamics
PAIN related	COMT	rs4680*; rs6269; Global MAF, G=0.37; rs 6433*; Global MAF, T=0.39; rs4818; Global MAF, G=0.32; rs207550; Global MAF, G=0.35	Val158Met + 3 more SNPs found in haplotype (see rs#)	Enzyme with a key role in catecholamine metabolism. Functional SNP(V158M) associated with increased sensitivity to painful stimuli (Zubieta et al. 2003) and need for lower doses of morphine in cancer patients (Cevoli et al. 2006). A haplotype of 4 SNPs of COMT associated with TMJ (Diatchenko 2005). *Previously genotyped in EAS
PAIN related	OPRM1 - Of secondary priority	rs1799971; Caucasian MAF, G=11-17%; African-American MAF, G=2.2%	A118G	μ -opioid receptor - binds endorphins, regulates pain signal transduction cascade. Patients with this variant have shown a lower pain threshold and a higher drug consumption in order to achieve the analgesic effect (Mura et al. 2013)
PAIN related	GCH1	rs3783641; Global MAF, A=0.23; rs8007267; Global MAF, T=0.28; rs10483639; Global MAF, C=0.29	Mulptiple SNPs and haplotypes	Guanosine triphosphate cyclohydrolase 1 the rate-limiting enzyme for tetrahydrobiopterin (BH4) synthesis acts as key modulator of peripheral neuropathic and inflammatory pain. BH4 is an essential cofactor for catecholamine, serotonin and nitric oxide production.
PAIN related	ADRB2	rs1042713; Global MAF, A=0.47; rs 1042714; Global MAF, G=0.47	Arg16Gly, Gln27Glu	beta-2-adrenergic receptor - member of the G protein-coupled receptor superfamily. Associated with sleep dysfunction in fibromyalgia, psychological distress
PAIN related	DBH	rs1611115*; Global MAF, T=0.21; rs1108580*; MAF G=0.43; rs6271; Global MAF, T=0.03; CH9	D159->T +1603C->T, Multiple SNPs	Dopamine beta hydroxylase - loss of function leads to increase in norepinephrine; pro-nociceptive - interaction with COMT? *Previously genotyped in EAS

ITEM	GENE	SNP ACCESSION #	OTHER CONVENTIONS	RATIONALE
PAIN related	5HTT	SLC6A4; 43% s/1 heterozygous; 18% s/s homozygous	44bp ins/del in promoter	5HT transporter – associated with chronic pain conditions, short allele associated with fibromyalgia
PAIN related	5HTR2A	HTR2A; Global MAF, A=0.43	102T>C SNP	5-hydroxytryptamine receptor 2A – related to increased risk of Fibromyalgia and TMD
PAIN related	SCN9A	Rs6746030; Global MAF, A=0.11	R1150W	Voltage-gated sodium channel which plays a significant role in nociception signaling. Associated with osteoarthritis, sciatica and post-amputation pain
PAIN related	IL6	rs1800795; MAF C=0.18		174G>C *Previously genotyped in EAS
PAIN related	ESR2beta	1730 G>A		
PAIN related	PANX1	Rs1138800 aa5; MAF 0.2883; Rs111535626; MAF NA aa152; Rs12793348; MAF 0.13 aa272; Rs74549886; MAF 0.05 aa390; Rs149967628 MAF 0.002 aa155; Rs148324299; MAF 0.00/1 aa 378		Member of the gap junction family of proteins, forms plasma membrane channels permeable to ATP and associated with P2X7 receptor; activates caspases-1 to the inflammasome. Abundantly expressed in CNS – glia and neurons and immune system (macrophages and T cells). Recently proposed to provide a link between CSD and headache in migraine.
Stress/mood	MICA	CH6; MICA*00801; European MAF=0.43; African-American MAF=0.27	HLA class I antigen; MHC class I chain-related gene A protein; MHC class I chain-related protein A; Stress inducible class I homolog	MHC class I polypeptide-related sequence A - functions as a stress-induced antigen; possible association of MICA*00801 heterozygotes with AD in subjects positive for the epsilon 4 allele of apolipoprotein E (Quiroga 2009)
Stress/mood	DR Of secondary priority	CH3; rs6280; Global MAF, G=0.25	D3DR; ETM1; FET1; MGC149204; MGC149205	Dopamine receptor D3 - encodes the D3 subtype of the five (D1-D5) dopamine receptors; receptor is localized to the limbic areas of the brain, which are associated with cognitive, emotional, and endocrine functions; glycine allele associated with paranoid and delusional ideations in AD (Sato 2009)
Stress/mood	ChAT	rs3810950; Global MAF, A=0.15; rs1880676; Global MAF, A=0.15	4G to A transition, others	Choline acetyltransferase (ChAT) - catalyzes the biosynthesis of acetylcholine. Polymorphisms associated with AD and MCI, depression and AD (Grunblatt 2009)
Stress/mood	BDNF	rs6265*; Global MAF, T=0.23; rs56164415; Global MAF, A=0.05; rs16917204; Global MAF, C=0.24	G196A (val66met), C270T, G11757 C	Brain derived neurotrophic factor - member of the nerve growth factor family; induced by cortical neurons, necessary for survival of striatal neurons. Expression is reduced in AD and HD. Postulated to play a role in the regulation of stress response and in the biology of mood disorders. *Previously genotyped in EAS
Stress/mood	Galanin (GAL)	rs948854; Global MAF, C=0.31		Encodes an estrogen-inducible neuropeptide, highly expressed in brain regions reported to be involved in regulation of mood and may have a direct modulatory effect on HPA regulation. SNP was associated with more severe anxiety symptoms and with higher HPA-axis activity at admission in females but not males (Unschuld et al., 2010).
Stress/mood	SERT, serotonin transporter	rs25531; Global MAF, C=0.11; rs25532; Global MAF, A=0.06	5-HTTLPR	Polymorphism results in either long (l) or short (s) 5HTT transcripts. The “l” allele leads to higher levels of 5HTT transcription than the “s”. Among individuals who experienced childhood trauma or recent stressful events, only s allele carriers show increased susceptibility to anxiety (Gunthert et al., 2007; Stein et al., 2008) and depressive symptoms (Caspi et al., 2003; Eley et al., 2004; Kaufman et al., 2004; Wilhelm et al., 2006; Zalsman et al., 2006). s allele is also associated with poorer delayed recall and, in combination with higher waking cortisol levels, predicted poorer memory and lower hippocampal volume in healthy, older adults (O’Hara et al. 2007).

ITEM	GENE	SNP ACCESSION #	OTHER CONVENTIONS	RATIONALE
Stress/mood	MAOA	rs1137070; MAF T=0.4		Oxidizes neurotransmitters and dietary amines. Low levels of MAO activity and mutations in the MAOA gene have been associated with violent, criminal, or impulsive behavior (Chen et al., 2004). Associated with antisocial behavior in children (Caspi et al. 2002) and restless leg syndrome (Desautels et al. 2002). *Previously genotyped in EAS
Autophagy	GAB-2	Rs10793294; Global MAF, C=0.49		GRB2-associated binding protein (GAB) gene family. act as adapters for transmitting various signals in response to stimuli through cytokine and growth factor receptors, and T- and B-cell antigen receptors. Possible modulator of Tau processing.
Autophagy	RELN	rs607755; Global MAF, G=0.46		Reelin - encodes a large secreted extracellular matrix protein thought to control cell-cell interactions critical for cell positioning and neuronal migration during brain development; expression decreased in AD
Autophagy	SORL1	rs661057; Global MAF, C=0.43; rs12364988; Global MAF, C=0.45; rs641120; Global MAF, A=0.42; CH11	Sorting protein-related receptor containing LDLR class A repeats; SorLA; Low-density lipoprotein receptor relative with 11 ligand-binding repeats; LR11	Sortilin related protein - encodes a mosaic protein that belongs to at least two families: the vacuolar protein sorting 10 (VPS10) domain-containing receptor family, and the low density lipoprotein receptor (LDLR) family; likely plays roles in endocytosis and sorting; genetic contributor to late-onset AD (LOAD) – appears to be through a female-specific mechanism
Sex Hormones	ESR1, estrogen receptor alpha	rs9340799; Global MAF, G=0.26	XBAL	Receptor is prevalent in brain regions that regulate memory and has been associated with depression, anxiety, verbal memory performance and risk of AD (See Sundermann et al., for review)
Sex Hormones	ESR1, estrogen receptor alpha	rs223493; Global MAF, G=0.29	PvuII	Receptor is prevalent in brain regions that regulate memory and has been associated with depression, anxiety, verbal memory performance and risk of AD (See Sundermann et al., for review)
Sex Hormones	ER beta	Rs4986938; MAF T=0.27		*Previously genotyped in EAS
Sex Hormones	CYP19, aromatase - Of secondary priority	rs767199; Global MAF, A=0.35		Encodes aromatase, the enzyme that converts androgens to estrogens. Associated with risk of AD independently (Butler et al., 2010) and in a 3 SNP haplotype (rs727479, rs1065778) (Iivonen et al., 2004) in men & women
Sex Hormones	CYP19, aromatase - Of secondary priority	rs1065778 Global MAF, C=0.37		Encodes aromatase, the enzyme that converts androgens to estrogens. Associated with risk of AD independently (Butler et al., 2010) and in a 3 SNP haplotype (rs727479, rs1065778) (Iivonen et al., 2004) in men & women
Sex Hormones	CYP19, aromatase - Of secondary priority	rs10046; Global MAF, A=0.41		Encodes aromatase, the enzyme that converts androgens to estrogens. Associated with risk of AD and MCI in women (Butler et al., 2010)
Sex Hormones	BCHE, butyrylcholinesterase	rs1803274; Global MAF, T=0.16; rs1126680; Global MAF, T=0.04	A539T, 116A	Encodes an enzyme that is upregulated in AD and is associated with decline in cholinergic activity in AD (Perry et al., 1978; Arendt et al., 1972; Furtado-Alle et al., 2008). Associated with formation of amyloid plaques, neurofibrillary tangles (Carson et al., 1991) and with risk of AD in women (Alvarez-Arcaya et al., 2000). Has been found to interact with CYP19 (Combarros et al., 2005), ESR1 and APOE in effect on AD risk.
Sex Hormones	TOMM40 / Of secondary priority	Rs8106922; Global MAF; G=0.3012		Encodes membrane subunit in outer core of mitochondria, AD risk, LOAD age at onset

ITEM	GENE	SNP ACCESSION #	OTHER CONVENTIONS	RATIONALE
Sex Hormones	MTHFR	rs1801131; Global MAF; G=0.22		Homocysteine pathway
Sex Hormones	PPAR	rs1800206; MAF G=0.0248		Altered DNA binding of PPARG, improved insulin sensitivity
Sex Hormones	APOC3	rs2542052*; MAF C=0.4839; upstream variant 2KB		Component of both high density lipoprotein (HDL) and apolipoprotein B (APOB; 107730)-containing lipoprotein particles, impairs catabolism and hepatic uptake of apoB-containing lipoproteins, appears to enhance the catabolism of HDL particles, enhances monocyte adhesion to vascular endothelial cells, and activates inflammatory signaling pathways. Associated with Longevity. *Previously genotyped in EAS
Sex Hormones	Adiponectin	rs17300539; MAF A=0.0399		Higher adiponectin, longevity, less CVD, metabolic syndrome
Sex Hormones	ADIPOR1	rs56354395; MAF A=0.48		The adiponectin receptors, ADIPOR1 and ADIPOR2, serve as receptors for globular and full-length adiponectin and mediate increased AMPK and PPAR-alpha ligand activities, as well as fatty acid oxidation and glucose uptake by adiponectin (Yamauchi et al., 2003).
Sex Hormones	FOXO3A	rs2764264; MAF C=0.4371; Rs13217795; MAF C=0.4040; Rs2802292; MAF G=0.449		Longevity in Hawaian heart (Japanese men)
Sex Hormones	CETP	rs708272; MAF A=0.3792; Rs 5882*; MAF G=0.44		The cholesteryl ester transfer protein mediates the exchange of lipids between lipoproteins, resulting in the net transfer of cholesteryl ester from high density lipoprotein (HDL) to other lipoproteins and in the subsequent uptake of cholesterol by hepatocytes. *Previously genotyped in EAS
PD/Dystonia related	LRRK2	rs34637584; (G2019S); MAF; A=0.0005; Rs1491942; Rs7133914		Associated with PD in Ashkenazi Jews (Ozelius LJ et al. 2006) Protective
PD/Dystonia related	BST1	Rs12502586; Rs4698412	(AJ), (European)	Bone marrow stromal antigen
PD/Dystonia related	MAPT	Rs2942168 (1st priority); Rs1052553; (2nd priority)		The microtubule-associated proteins tau coassemble with tubulin into microtubules in vitro. enriched in axons.
PD/Dystonia related	SNCA	Rs356220 (1st priority); Rs181489		Alpha-synuclein is a highly conserved protein that is abundant in neurons, especially presynaptic terminals. Aggregated alpha-synuclein proteins form brain lesions that are hallmarks of neurodegenerative synucleinopathies (summary by Giasson et al., 2000).
PD/Dystonia related	TARDBP	Rs11689432	A382T	The TARDBP gene encodes the 43-kD TAR DNA-binding protein, which was originally identified as a transcriptional repressor that binds to TAR DNA of human immunodeficiency virus type 1. It is also involved in regulation of gene expression and splicing (summary by Benajiba et al., 2009). Associated with ALS and FTD.
PD/Dystonia related	TMEM106B	rs 1990622		Homozygous T leads to problems TMEM106B SNPs (van Deerlin)
PD/Dystonia related	Progranulin – Granulin precursor GRN - Priority	rs5848; rs63751294		88-kD glycoprotein that functions as an autocrine growth factor – involved in FTD and Neuronal Ceroid Lipofuscinosis

ITEM	GENE	SNP ACCESSION #	OTHER CONVENTIONS	RATIONALE
PD/Dystonia related	GBA	7 SNPs.; Rs104886460; Rs387906315; Rs76763715; Rs 421016; Rs80356769; Rs1064651; Rs2230288		Acid beta-glucocerebrosidase, also known as beta-glucosidase (GBA) is a lysosomal enzyme that catalyzes the breakdown of the glycolipid glucosylceramide (GlcCer) to ceramide and glucose (Beutler, 1992). Risk for diffuse Lewy body disease, late onset PD and Gauchers.
PD/Dystonia related	STK39	Rs2102808		SERINE/THREONINE PROTEIN KINASE 39
PD/Dystonia related	MCCC1/LAMP3	Rs11711441		encodes the alpha subunit of 3-methylcrotonyl-CoA carboxylase, a biotin-dependent mitochondrial enzyme essential for the catabolism of leucine.
PD/Dystonia related	GPNMB	Rs156429		Glycoprotein NMB – marker of melanocyte tumor progression evolves
PD/Dystonia related	RIT2/SYt2	Rs12456492		RIC-LIKE PROTEIN WITHOUT CAAX MOTIF 2 – Ras family of small GTPases – expressed in neurons

4.2 MCI Diagnoses

Return to Section [2.2.3](#)

Final Primary Diagnosis

Code	Description
0	Normal
1	Probable Alzheimer's Disease
2	Possible Alzheimer's Disease
3	Probable Ischemic Vascular Dementia
4	Possible Ischemic Vascular Dementia
5	Binswanger's Syndrome
6	Possible/Probable Dementia with Lewy Bodies
7	Frontotemporal Dementia
8	Symptomatic hydrocephalus
9	Hypothyroidism
10	Subacute combined degeneration – B12
11	Traumatic brain damage
12	Neurosyphilis
13	AIDS dementia
14	Brain tumor
15	Creutzfeldt Jacob disease
16	Down's syndrome
17	Herpes encephalitis
18	Huntington's disease
19	Leukodystrophy (specify type)
20	Motor neuron disease
21	Multiple sclerosis
22	Multi-system atrophy
23	Progressive subcortical gliosis
24	Progressive supranuclear palsy
25	Mixed Dementia – Alzheimer's + Vascular
26	Parkinsonian Dementia
27	Other types
28	Pure amnesic syndrome
29	Alcoholic encephalopathy
30	Corticobasal degeneration
31	Memory Impairment
32	Memory Impairment & Functional Decline
33	Cognitive Impairment, No Memory Impairment
34	Cognitive Impairment & Functional Decline
41	Dementia (indeterminate)
42	Memory Impairment & Cognitive Impairment, No Functional Decline
50	Major depression (DSM-IV)
56	Parkinson's disease (without dementia)
59	Hyperthyroidism
60	aMCI
99	Insufficient Information