ADSP Phenotype Harmonization Consortium –

Derivation of Harmonized Fluid Biomarker Levels

Oscar Diaz Ruiz1,2, Victoria Fernandez Hernandez1,2, John Budde1,2, Pat Kohlfeld1,2,3, Joanna Norton1,2,3, Jen Gentsch1,2,3, Jigyasha Timsina1,2, Kuheli Sen1,2, Ellen Liu1,2, Derek B. Archer5,6,9, Logan C. Dumitrescu5,6,9, Michael L. Cuccaro7,8, Timothy J. Hohman 5,6,9, Carlos Cruchaga,1,2,3

1  Neurogenomics and Informatic Center, Washington University School of Medicine in St. Louis, St. Louis, MO,   
 USA.

2  Department of Psychiatry, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.

3  Knight Alzheimer Disease Research Center, Washington University School of Medicine in St. Louis, St. Louis,   
 MO, USA.

4  Hope Center for Neurological Disorders, Washington University School of Medicine in St. Louis, St. Louis, MO,   
 USA.

5  Vanderbilt Memory and Alzheimer’s Center, Vanderbilt University Medical Center,

Nashville, TN, USA.

6 Vanderbilt Genetics Institute, Department of Medicine, Vanderbilt University Medical Center,

Nashville, TN, USA.

7  The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA.

8  Dr. John T. MacDonald Foundation, Department of Human Genetics, University of Miami,

Miami, FL, USA.

9 Dept. of Neurology, Vanderbilt University Medical Center, Nashville, USA.

|  |  |
| --- | --- |
| **Contents** | |
| Page 1 | Summary |
| Page 1 | 1. Preliminary analyses of each study included in the legacy co-calibration model |
| Page 1 | Step A1. Acquire data and documentation from each study |
| Page 1 | Step A2. Choice of base dataset |
| Page 1 | Step A3. Data quality control |
| Page 2 | 2. Co-calibration of the biomarker levels across ADNI, NACC, and Knight ADRC |
| Page 2 | Step B1: Z-score calculations |
| Page 3 | References |
| Page 3 | Version Information |
| Page 4 | About the Authors |

**Summary**

Platforms that measure fluid biomarker levels differ across studies and cohorts, complicating the ability to merge biomarker data across cohorts. Co-calibrated scores generated using modern statistical approaches permit direct comparison of study participants in different studies who were assessed with different biomarker platforms. Fluid biomarker scores were harmonized across the following datasets within the ADSP: Alzheimer’s Disease Neuroimaging Initiative (ADNI), National Alzheimer’s Coordinating Center (NACC), and the Memory and Aging Project at Knight Alzheimer’s Disease Research Center (MAP at Knight ADRC). The co-calibrated scores are standardized on the same metric, making the harmonized scores directly comparable even across studies that administered different tests.

**1. Preliminary analyses in each study included in the Legacy co-calibration model**

“Legacy” here refers to the first set of studies we evaluated—ADNI, NACC, and Knight ADRC. ADNI and NACC are two large, widely used studies with at least moderate sized fluid biomarker levels.

## STEP A1: Acquire data and documentation from each study. We establish data use agreements for each study and acquire granular level data from fluid biomarker levels along with detailed documentation on each of the items included in the measurements. Information that has proven to be useful includes method of data collection, description of platform used for analysis, specific stimuli administered, and information on how levels are coded. Data dictionaries and protocol documents have proven to be useful. This step might take multiple iterations as we learn more about the data set.

## **STEP A2: Choice of Base Dataset*.*** In each of the studies (Alzheimer’s Disease Neuroimaging Initiative [ADNI], the National Alzheimer’s Coordinating Center [NACC], and the Memory and Aging Project at Knight Alzheimer’s Disease Research Center [Knight ADRC]), varying numbers of participants had fluid biomarker levels drawn, and even less had individuals with multiple draws where simple quality control trends analysis could be observed. As a result, the most informative dataset, in this case ADNI, was defined as by the maximum number of unique data points available and was consequently used for downstream analysis.

STEP A3: Data Quality Control. We obtained access to each fluid biomarker datasets andperformed initial quality control steps on the data.These steps included the removal of any duplicate data points, as defined by having the same sample ID and CSF draw date.Missing data were a particular area of focus. Each study had very little information beyond the fact of a missing data element. It was unknown whether the missing data was due to a sample collection error, sample processing error, or participant refusal. Because we could not decipher the distinction, we ended up treating all types of missing data as if that biomarker was not collected. The last measure of quality control was the removal of outlier biomarker levels. Biomarker level outliers were calculated using the log10 transformed biomarker values. They were then flagged as outliers if they were found outside the Q3 + 1.5 IQR and Q1 – 1.5 IQR (where Q1 and Q3 represent the first and third quartile range respectively, and IQR represents the interquartile range). Like with missingness, we were unable to determine the cause of the outlying data and deemed it unreliable.

**2. Co-calibration of the biomarker levels across ADNI, NACC, and Knight ADRC**

**STEP B1: Z-score Calculations.** Z-scores were calculated for each of the following harmonized fluid biomarker levels: Amyloid-beta, Tau, and pTau. They were computed by scaling all datapoints that passed through quality control, to have a mean of 0 and a variance of 1 using the “scale” function in R statistical software (v3.5.2).

# References­

1. Corwin EJ, Moore SM, Plotsky A, Heitkemper MM, Dorsey SG, Waldrop-Valverde D, Bailey DE, Docherty SL, Whitney JD, Musil CM, Dougherty CM, Mccloskey DJ, Austin JK, Grady PA, University E, Hodgson Woodruff N (2017) Feasibility of Combining Common Data Elements Across Studies to Test a Hypothesis. *Journal of Nursing Scholarship* 49, 249–258.

2. Basu A, Warzel D, Eftekhari A, Kirby JS, Freymann J, Knable J, Sharma A, Jacobs P (2019) Call for Data Standardization: Lessons Learned and Recommendations in an Imaging Study. *JCO Clin Cancer Inform* 1–11.

3. Ebenau JL, Timmers T, Wesselman LMP, Verberk IMW, Verfaillie SCJ, Slot RER, van Harten AC, Teunissen CE, Barkhof F, van den Bosch KA, van Leeuwenstijn M, Tomassen J, den Braber A, Visser PJ, Prins ND, Sikkes SAM, Scheltens P, van Berckel BNM, van der Flier WM (2020) ATN classification and clinical progression in subjective cognitive decline The SCIENCe project.

4. Deming Y, Li Z, Kapoor M, Harari O, Del-Aguila JL, Black K, Carrell D, Cai Y, Fernandez MV, Budde J, Ma S, Saef B, Howells B, Huang K lin, Bertelsen S, Fagan AM, Holtzman DM, Morris JC, Kim S, Saykin AJ, de Jager PL, Albert M, Moghekar A, O’Brien R, Riemenschneider M, Petersen RC, Blennow K, Zetterberg H, Minthon L, van Deerlin VM, Lee VMY, Shaw LM, Trojanowski JQ, Schellenberg G, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Peskind ER, Li G, di Narzo AF, Kauwe JSK, Goate AM, Cruchaga C (2017) Genome-wide association study identifies four novel loci associated with Alzheimer’s endophenotypes and disease modifiers. *Acta Neuropathol* 133, 839–856.

5. Li Y, Ammari S, Balleyguier C, Lassau N, Chouzenoux E (2021) Impact of preprocessing and harmonization methods on the removal of scanner effects in brain mri radiomic features. *Cancers (Basel)* 13.

**Version Information**

This is the first version of this document.

Dataset Information

This methods document applies to the following dataset(s) available from the ADNI repository:

|  |  |
| --- | --- |
| **Dataset Name** | **Date Submitted** |
| ADSP Phenotype Harmonization Consortium (PHC) - Composite Biomarker Scores | 8 December 2023 |

# About the Authors

This document was prepared by Dr. Cruchaga and Dr. Diaz Ruiz from Washington University, along with Drs. Hohman, Dumitrescu, and Archer from Vanderbilt University Medical Center. For more information, please contact Dr. Hohman by email at [timothy.j.hohman@vumc.org](mailto:timothy.j.hohman@vumc.org).

*Notice: This document is presented by the author(s) as a service to ADNI data users.  However, users should be aware that no formal review process has vetted this document and that ADNI cannot guarantee the accuracy or utility of this document.*