HHS Public Access

Author manuscript

J Clin Epidemiol. Author manuscript; available in PMC 2026 February 01.

Published in final edited form as:

J Clin Epidemiol. 2025 February ; 178: 111620. doi:10.1016/j.jclinepi.2024.111620.

Development of a refined harmonization approach for longitudinal cognitive data in people with HIV

Lang Lang[a,†], **Leah H. Rubin**[b,c,d,e,†,**], **Raha M. Dastgheyb**[b], **David E. Vance**[f], **Scott L. Letendre**[g], **Donald R. Franklin Jr**[g], **Yanxun Xu**[a,h,†,**]

[a]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA

[b]Departments of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[c]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[d]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

[e]Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[f]School of Nursing, University of Alabama at Birmingham, Birmingham, AL, USA

[g]HIV Neurobehavioral Research Program, Departments of Medicine and Psychiatry, University of California San Diego, CA, USA

[h]Division of Biostatistics and Bioinformatics at The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Abstract

Objective: The aim of this study was to develop a refined method for harmonizing longitudinal cognitive data across several large-scale studies in people with HIV (PWH), in whom cognitive complications are common and heterogeneous in presentation.

Study Design and Setting: We developed a refined method for harmonizing longitudinal cognitive data across five large-scale studies in PWH that used different cognitive batteries with only some overlapping tests—Women's Interagency HIV Study (WIHS), Multicenter

Requests for reprints and correspondences should be addressed to either: Yanxun Xu, Ph.D., Department of Applied Mathematics and Statistics, Johns Hopkins University Whiting School of Engineering, 3400 North Charles Street, Wyman N429, Baltimore, MD 21218, yanxun.xu@jhu.edu, Leah H. Rubin, Ph.D., MPH, Department of Neurology, Johns Hopkins University School of Medicine, 600 N. Wolfe Street/ Carnegie 3-301, Baltimore, MD. 21287-7613, lrubin@jhmi.edu.

†equally contributing

**co-corresponding

Declaration of interests

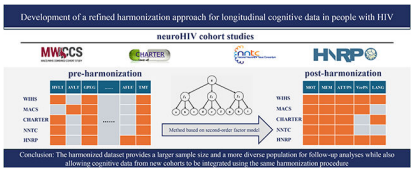
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AIDS Cohort Study (MACS), CNS HIV Antiretroviral Therapy Effects Research (CHARTER), National NeuroAIDS Tissue Consortium (NNTC), and the HIV Neurobehavioral Research Program (HNRP). Traditional data harmonization methods using latent variable models focus on crosssectional data and require the presence of common cognitive tests to serve as “linking” assessments. However, the absence of such common tests for certain cognitive domains can preclude the direct application of these traditional techniques. To address these challenges, we developed a harmonization method that leveraged a second-order factor model, which capitalized on the structural relationships among cognitive domains.

Results: Our approach yielded harmonized cognitive domain scores that are demographically consistent across different cohorts and exhibit strong correlations with the raw or log transformed (e.g., timed outcomes) cognitive test scores. These harmonized scores accurately reflected variations according to age, educational status, and other demographic factors, while preserving participants longitudinal cognitive trajectories.

Conclusion: Our harmonization methods are essential for future analyses of large-scale, retrospective data to understand the heterogeneity in cognitive complications in PWH. These methods can be applied to harmonize new datasets with similar measures.

Graphical Abstract



Plain Language Summary

Background: People with HIV (PWH) often experience cognitive complications, which can vary widely from person to person. To better understand individual differences in cognitive function among PWH, we need to combine and analyze large sets of cognitive data from various studies. However, this idea of combining data to study cognitive function in PWH has not been fully explored.

What We Did: We created a refined method to combine, or “harmonize,” cognitive test results from five large studies involving PWH. These studies used different sets of cognitive tests, and sometimes didn’t have any tests in common for certain mental abilities. Traditional methods require common tests to link the data together, so we developed a refined approach that uses the relationships between different cognitive abilities to align the data across studies.

What We Found: Our method successfully produced consistent scores for various cognitive domains across studies. These harmonized scores accurately reflected differences based on age, education, and other demographic factors. Importantly, they also preserved how each person’s cognitive abilities changed over time.

Why It Matters: Our refined harmonization method enables researchers to combine data from multiple studies even when common cognitive tests are lacking — a limitation of traditional methods. By harmonizing these datasets, we leverage larger sample sizes and increased diversity from merged cohorts. This enhances understanding of cognitive complications in people living

with HIV, potentially leading to better treatments and support. Additionally, our method can be used to harmonize new datasets that have similar measures but use different tests.

Keywords

Cognition; Factor model; Harmonization; HIV; Psychometrics

1 Introduction

Cognitive complications are common among people with HIV (PWH). The underlying pathophysiology of these complications are complex, and effective therapeutics are lacking. Advances in the field have been limited by the substantial heterogeneity in clinical presentation and the multifactorial nature of the underlying pathophysiology. This is in part influenced by the high prevalence of comorbid conditions (e.g., substance use, coinfections, cardiovascular and metabolic disease), antiretroviral therapy (ART) and non-ART drug use, and social determinants of health. Although numerous studies have identified cognitive profiles in smaller subgroups (e.g., virally suppressed [VS] PWH, PWH starting ART, VS-women with HIV [1] [2] [3] [4] [5] [6]), no single cohort of longitudinally followed PWH exists with cognitive data that adequately samples PWH from different geographic locations who vary in terms of the factors contributing to heterogeneity in cognition in PWH. Determining unbiased cognitive phenotypes using existing data and advancing our understanding of the pathophysiology of these complications requires harmonization of large-scale, longitudinal cognitive data from multiple cohorts. However, the application of harmonization for cognitive data among PWH remains relatively unexplored.

Harmonizing cognitive data outside of the neuroHIV field has proven effective for facilitating direct comparisons of cognitive outcomes across diverse cultural and demographic populations in the context of aging and other neurodegenerative conditions [7]. To date, a variety of data harmonization methods have been employed in these populations [8]. Latent variable models have been favored because they effectively accommodate variations across studies, with well-developed procedures for the linear factor model and widely-used item response theory models [9]. However, traditional data harmonization methods based on latent variable models depend on the presence of common items to assess the same latent trait [10], a condition often unmet due to minimally overlapping cognitive test batteries administered across cohorts. This challenge is not uncommon in HIV studies. For example, the Women's Interagency HIV Study (WIHS) used the Hopkins Verbal Learning Test-Revised (HVLT-R) to assess verbal memory whereas the Multicenter AIDS Cohort Study (MACS) used the Rey Auditory Verbal Learning Test (RAVLT). Each assessment uses different word lists: HVLT-R has 12 nouns with four words drawn from each of three semantic categories, whereas RAVLT has 15 unrelated nouns. There are no common items for harmonizing memory function [11]. Additionally, variations in test administration can render scores incomparable. For example, the Grooved Pegboard test is widely used across HIV cohorts, but some cohorts implemented a four-minute stopping rule for incomplete tasks, while others used a five-minute rule. Finally, traditional literature on harmonizing cognitive assessments primarily focuses on cross-sectional data and there is no widely agreed-upon procedure for longitudinal data.

We developed a refined method to harmonize longitudinal cognitive data that lack common tests among some cognitive domains by leveraging the latent structural information among cognitive domains. This refined approach adapts to the constraints of available data and ensures that cognitive trajectories are consistently tracked over time, providing a more accurate and comprehensive understanding of cognitive change in PWH.

2. Methods

2.1 Study participants

Cognitive data was harmonized across five HIV cohorts with longitudinal data (assessed every 6 months to 2 years). Participants' data were included in the analysis if they had completed at least one cognitive test battery and had sociodemographic data available (age, biological sex, race/ethnicity, education).

1. WIHS (https://statepi.jhsph.edu/wihs/wordpress/) [12] [13] [14]. Data were collected from April 2009-March 2019 (N=2637) and included only women with (n=1809) or without HIV (n=828).
2. MACS (https://statepi.jhsph.edu/macs/macs.html) [15]. Data were collected from February 1991-October 2019 (N=3635) and included only men who have sex with men, with (n=2078) or without HIV (n=1557).
3. CNS HIV Antiretroviral Therapy Effects Research (CHARTER; https://www.nntc.org/content/relationship-charter). Data were collected from February 2002-January 2020 (N=1528) and included primarily men (77%) with HIV.
4. National NeuroAIDS Tissue Consortium (NNTC; https://nntc.org/) [16]. Data were collected from February 1999-July 2020 (N=1321) and included almost all men and women with HIV (n=1316).
5. HIV Neurobehavioral Research Program (HNRP; https://hnrp.hivresearch.ucsd.edu/). Data were from various NIH-funded research studies at the University of California, San Diego's HNRP. Data were collected from May 1999-March 2020 (N=2024) and included primarily men (79%) with (n=1216) or without HIV (n=808).

See Supplemental Materials for details on each cohort. We recognize that registering secondary analyses is considered the best practice to promote transparency and reproducibility. However, we did not register our study because the datasets used in this analysis are well-established, and we do not have the necessary permissions to share them. Table 1 provides demographic characteristics of participants (from first visit only) contributing data from each cohort.

2.2 Cognitive assessments

Table 2 provides the cognitive tests and outcomes (raw or log transformed and reverse coded for timed data) considered for analysis, along with their respective presence in each cohort.

2.3 Statistical analyses

As an overview, we initially extracted data from the participants' first visit in each cohort, termed *reference line data*. Exploratory factor analysis (EFA) was performed for the cognitive tests within each cohort. EFA was selected to identify cognitive domains that accurately reflect the observed data structure [17]. This approach is particularly important, as neuroHIV studies often reveal discrepancies between the empirical structure observed in actual data and the theoretical structure described in the literature [18]. Guided by the EFA results, we used a second-order linear factor model to address the challenge posed by the absence of common tests for certain factors in some cohorts. We also verified the adequacy of the proposed factor structure against the cohort's reference line data before moving into the harmonization procedure. WIHS contained all five factors of interest and had a larger sample size compared to the other cohort with the five factors (HNRP). Thus, we used WIHS as the reference cohort, harmonizing the reference line data across cohorts by fixing parameters common in the second-order linear factor model. This step is also known as fixed parameter calibration [19]. With each cohort's reference line data harmonized to the same metric, we next applied the derived parameters from each cohort's reference line model to their full longitudinal data, allowing us to map the longitudinal data onto the reference line scale and provide a harmonized longitudinal cognition score for each cohort. Each step is detailed below. For reporting guidelines [20], see supplemental section "STROBE Statement".

2.3.1 Determining first-order factor structure—For the cognitive tests considered, EFA was initially performed within each cohort. We employed the following criteria to select the appropriate tests and establish the first-order factor structure across all cohorts for subsequent harmonization processes.

- Bayesian Information Criterion (BIC): Models were selected that exhibited lower BIC values, suggesting better model efficiency in terms of fit versus complexity.
- Consistency Across Cohorts: A uniform factor structure for common tests was favored to ensure the underlying constructs were measured equivalently across cohorts.
- Exclusion Criteria: Tests exhibiting disparate primary factor structures across cohorts or those with small loadings (<0.4) were excluded to maintain robust factor structures.
- Sufficient Test Representation: Each factor needs to be supported by at least two tests demonstrating high loadings for identifiability.

We then conducted Confirmatory Factor Analysis (CFA) to confirm the EFA factor structure within each cohort by examining the following fit statistics: Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). These statistics provide quantitative metrics on model fitting. Generally, CFI > 0.95, RMSEA < 0.05, and SRMR < 0.05 indicates a good fit, while a RMSEA < 0.08, and SRMR < 0.08 indicates an acceptable fit [21] [22] [23].

2.3.2 Verifying the second-order linear factor model—We fitted a second-order linear factor model to each cohort's reference line data, using the first-order factor structure identified via EFA and CFA. This model incorporated a general function factor represented by all first-order domain factors, capturing the underlying cognitive constructs across assessments. To assess the model accuracy and adequacy, we employed the same key fit indices as for the first-order structure (CFI, RMSEA, SRMR).

2.3.3 Harmonization of reference line data—Under the general framework of fixed parameter calibration [19], we harmonized reference line data across cohorts, starting with WIHS data using the proposed factor model under regular identification conditions on the factor variance and mean [24]. We then sequentially adapted this model for other cohorts fixing intercepts, factor loadings, and residual variances from previous cohort fittings, while allowing free estimation of factor variances, means, and parameters for new tests that were not included in previous cohorts. This step produced model parameters on a common scale within and across cohort's reference line data.

A key challenge in linking domains without common items is placing the means of these domain factors on the same scale. We address this by using a second-order factor model, linking the means of domain-specific factors (first-order factors) through their connection to the general cognitive function factor. Specifically, for cohorts other than WIHS (where the mean of the general factor is fixed at 0 for identification), we set the means of the first-order domain factors to zero and allow the general factor's mean to be freely estimated. This approach ensures that the domain factor means are functions of the general factor's mean and their respective loadings. If a cohort has a domain without common items shared with others, the general factor's mean can still be placed on the same scale as the other cohorts since it shares linking items from other domains, thereby also aligning the domain without linking items to the same scale as well. The major assumption for this approach is that the mean structures between the general function factor and the first-order factors should not be highly divergent across the cohorts being harmonized.

2.3.4 Harmonization of longitudinal data—Once the reference line data across cohorts were aligned to the same metric, we extended the harmonization to the full longitudinal data of each cohort. We fit the proposed factor models to each cohort's complete data set, fixing parameters to those from the reference line phase and allowing only factor means and variances to vary. We also calculated factor scores for each visit, providing harmonized longitudinal cognitive scores across cohorts. Figure 1 provides a schematic of this process. EFA was conducted using psych package [25] in R [26], while the more complex factor models were fitted using a robust maximum likelihood estimator in Mplus version 8.3 [27]. Mplus syntax used for reference line data harmonization and longitudinal harmonization is available at https://github.com/BHPDataSci/DataHarmonization.

3. Results

3.1 Cohort demographics for the reference line visit

Participants included 11,145 (7947 PWH; 3198 people without HIV [PWoH]), aged 18 to 99 years-old, with 36.5% women, 38.8% Black and 15.4% Hispanic, and 79.7% completed

≥12 years of education. Across cohorts, participants differed in terms of age, education, race/ethnicity, and proportion of PWH and visits completed (Table 1).

3.2 First-order factor structure obtained through EFA

Figure 2 illustrates the cognitive assessments included in our harmonization process and the first-order factor structure determined by EFA in 2.3.1 (Supplemental Tables 1–5). The structure aligned closely with standard cognitive domains and showed good performance in first-order CFA (CFI > 0.95, RMSEA ~ 0.06, SRMR < 0.05 for all cohorts, Supplemental Table 6 for detailed fit statistics). Five factors emerged across cohorts and were defined by the cognitive test outcomes demonstrating the highest factor loadings: *Factor 1-Declarative Memory* (verbal learning and memory test outcomes), *Factor 2-Verbal Processing Speed* (Stroop), *Factor 3-Attention/Processing Speed* (Trail Making Test, Paced Auditory Serial Addition Test), *Factor 4-Motor Function* (Grooved Pegboard Test), and *Factor 5-Language* (animal, letter, and action fluency). WAIS-III Digit Symbol test and Wisconsin Card Sorting Test were excluded from the EFA due to low factor loadings and inconsistent factor structures across cohorts.

3.3 Diagnostic statistics showed good fitting performance of second-order factor model to reference line data

The model fitting results indicate that our models performed well across all cohorts, with CFI values > 0.95, RMSEA ~ 0.06, and SRMR < 0.05. Details are provided in Supplemental Figures 1–5 and Supplemental Table 7.

3.4 Raw cognitive test scores associate with harmonized factor scores

To evaluate the effectiveness of our harmonization process, we examined associations between the raw test scores (note in the case of timed tests, these outcomes were log transformed and then reverse scored) and their harmonized equivalents across cohorts. All correlations were significant (P's<0.001) and > 0.7 between the raw or log transformed test score and their harmonized equivalents across cohorts (details in Supplemental Table 8 (a) and Supplemental Figure 6). Figure 3B provides an illustration of the link between Grooved Pegboard test outcomes and *Factor 5-Motor Function* by cohort. The relative positions and shapes of the log transformed/reverse scored test scores were preserved in the harmonized motor factor scores, demonstrating that the harmonization process maintains reasonable alignment with the raw scores.

3.5 Harmonized scores maintain expected differences across demographic variables

Next, we examined associations between sociodemographic factors (education, age, sex, race/ethnicity, HIV status) and raw scores followed by associations with harmonized scores. The pattern of associations between these factors with raw and harmonized scores were similar. Trends and significant differences across demographics were consistent between raw and harmonized factor scores (Supplemental Tables 11–18; Supplemental Figures 7–10).

3.6 Harmonized scores preserve longitudinal trends for individuals

We analyzed the longitudinal harmonized factor scores to determine if they preserved the trends of individuals performance over time, demonstrating the effectiveness of the harmonization process. Figure 3C presents the longitudinal harmonized motor factor scores for participants with more than 33 visits in the MACS cohort (n = 25). These participants were selected for illustration because they had numerous visits across a relatively long time interval to observe the cognitive trajectories. These trajectories highlight a general trend of decline in motor skills as age increases which is consistent with existing research [28].

3.7 Distributions of harmonized cognition factor scores by cohorts

After applying the harmonization procedure in section 2.3.3 and 2.3.4, we obtained harmonized reference line and longitudinal cognition factor scores. Figure 3A shows the box plots of harmonized domain scores from each cohort. Each box plot represents the distribution for a specific cognitive domain, with different colors denoting each cohort. As summarized in Table 3, MACS and HNRP demonstrated better performance across all cognitive factors whereas NNTC demonstrated the lowest performance across all cognitive factors.

4. Discussion

We developed a refined method to harmonize cognitive data across five large-scale, longitudinal HIV cohort studies, particularly when the test batteries measuring these domains differed between cohorts. The harmonized scores for these cognitive domains showed reasonable distributions aligned with the demographic characteristics across cohorts, high correlation with the raw/log transformed test scores, and expected differentiation according to education and age. Additionally, these scores preserved longitudinal trajectories of individual cognitive changes. Overall, the validation analysis confirmed that our harmonization method produced a robust and reasonable harmonization of scores. For integration of new data, see supplemental section “Integration of new data”.

To date, various data harmonization techniques have been applied [8], the most commonly used are standardization methods like converting performances to a common scale using T-scores or Z-scores. However, these methods assume similar distributions across populations and may not account for group differences. Less frequently used are multiple imputation techniques (e.g., multiple imputation with chained equations, random forests), which assume cognitive tests lacking direct linking are missing at random. A key debate is whether intentionally missing variables (missing by design) can be accurately imputed, as their missingness probability is typically one. The final category involves latent variable modeling. Traditional approaches start with unidimensional latent variable models and use fixed parameter calibration or concurrent calibration but often struggle due to limited overlap of cognitive tests across cohorts. Our proposed method also fits into this category but addresses the challenge of no common items in certain domains by introducing a second-order factor model. This model links domains through a shared general factor, enabling coherent scaling even without direct linking items.

For our method to function effectively and yield meaningful results, two key assumptions must be satisfied. First, the second-order factor model needs to fit the data adequately. Second, the mean structures between the general function factor and the first-order factors should not be highly divergent across the cohorts being harmonized. See supplemental section “Detailed Explanation of Assumptions and Limitations” for more detailed discussion. A recent study [29] proposed to linearly link domain factors with few or no common tests to a subsuming trait (e.g., general cognitive performance) where sufficient information exists for harmonization. This approach required standard item response theory assumptions—unidimensionality, local independence, and suitable item response functions—as well as stringent conditions for linear equating [30], matching relative distributions between the subdomain and the subsuming trait, and the assumption that linked domains are vertical reflections of the reified domains. In contrast, our method avoids the latter three assumptions by leveraging a second-order factor model, though it introduces assumptions about the model structure and mean consistency across cohorts. Both methods have their applicable scenarios, and future studies should carefully assess which assumptions are suitable for their specific context. When appropriate, results from both methods can be used to cross-validate each other. A connection can also be found between our proposed method and Nonequivalent Groups with Covariates (NEC) approach [31] [32]. A detailed discussion of this connection is provided in the supplemental section “Connection with NEC”.

When interpreting the harmonized results, it is important to recognize that the factor scores derived for latent variables not only leverage correlations among tests within their own domains but also incorporate correlations with tests from other domains. This is due to the nature of the second-order factor model, where first-order factors (cognitive domains) are correlated with each other through the second-order (general cognitive function) factor, along with the correlations observed among all cognitive tests in real data. Given the potential for cross-domain influence, future studies should be mindful of this overlap. Furthermore, the precision of factor scores depends on the number of tests used to measure a domain. While fewer tests may reduce measurement bias, this comes at the cost of reduced precision. Future studies should carefully consider this trade-off when designing domain-specific assessments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The authors gratefully acknowledge the contributions of the study participants for all studies (MWCCS, CHARTER, NNTC, HNRP) and dedication of the staff at the sites.

Funding:

The contents of this publication are solely the responsibility of the authors and do not represent the official views of the NNTC or National Institutes of Health (NIH). The contents of this publication are solely the responsibility of the authors and do not represent the official views of the National Institutes of Health (NIH). MWCCS (Principal Investigators): Atlanta CRS (Ighovwerha Ofotokun, Anandi Sheth, and Gina Wingood), U01-HL146241; Baltimore CRS (Todd Brown and Joseph Margolick), U01-HL146201; Bronx CRS (Kathryn Anastos, David Hanna, and Anjali Sharma), U01-HL146204; Brooklyn CRS (Deborah Gustafson and Tracey Wilson), U01-HL146202; Data Analysis and Coordination Center (Gypsyamber D’Souza, Stephen Gange and Elizabeth

Topper), U01-HL146193; Chicago-Cook County CRS (Mardge Cohen, Audrey French, and Ryan Ross), U01-HL146245; Chicago-Northwestern CRS (Steven Wolinsky, Frank Palella, and Valentina Stosor), U01-HL146240; Northern California CRS (Bradley Aouizerat, Jennifer Price, and Phyllis Tien), U01-HL146242; Los Angeles CRS (Roger Detels and Matthew Mimiaga), U01-HL146333; Metropolitan Washington CRS (Seble Kassaye and Daniel Merenstein), U01-HL146205; Miami CRS (Maria Alcaide, Margaret Fischl, and Deborah Jones), U01-HL146203; Pittsburgh CRS (Jeremy Martinson and Charles Rinaldo), U01-HL146208; UAB-MS CRS (Mirjam-Colette Kempf, James B. Brock, Emily Levitan, and Deborah Konkle-Parker), U01-HL146192; UNC CRS (M. Bradley Drummond and Michelle Floris-Moore), U01-HL146194. The MWCCS is funded primarily by the National Heart, Lung, and Blood Institute (NHLBI), with additional co-funding from the *Eunice Kennedy Shriver* National Institute Of Child Health & Human Development (NICHD), National Institute On Aging (NIA), National Institute Of Dental & Craniofacial Research (NIDCR), National Institute Of Allergy And Infectious Diseases (NIAID), National Institute Of Neurological Disorders And Stroke (NINDS), National Institute Of Mental Health (NIMH), National Institute On Drug Abuse (NIDA), National Institute Of Nursing Research (NINR), National Cancer Institute (NCI), National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Deafness and Other Communication Disorders (NIDCD), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute on Minority Health and Health Disparities (NIMHD), and in coordination and alignment with the research priorities of the National Institutes of Health, Office of AIDS Research (OAR). MWCCS data collection is also supported by UL1-TR000004 (UCSF CTSA), UL1-TR003098 (JHU ICTR), UL1-TR001881 (UCLA CTSI), P30-AI-050409 (Atlanta CFAR), P30-AI-073961 (Miami CFAR), P30-AI-050410 (UNC CFAR), P30-AI-027767 (UAB CFAR), P30-MH-116867 (Miami CHARM), UL1-TR001409 (DC CTSA), KL2-TR001432 (DC CTSA), and TL1-TR001431 (DC CTSA). This publication was also made possible by the NNTC project which is a funded contract mechanism supported by NIMH, NIDA, NIA, and NINDS: Manhattan HIV Brain Bank (MHBB): U24MH100931; Texas NeuroAIDS Research Center (TNRC): U24MH100930; National Neurological AIDS Bank (NNAB): U24MH100929; California NeuroAIDS Tissue Network (CNTN): U24MH100928; Data Coordinating Center (DCC): U24MH100925. This work was in part supported by the Johns Hopkins Center for the Advancement of HIV Neurotherapeutics (P30MH075773; Rubin, Slusher, Clements), CHARTER (R24 MH129166; Letendre, Ellis), the HIV Neurobehavioral Research Center (P30 MH062512; Moore, Ellis), National Science Foundation grants 1918854 (Xu), and R01 MH128085 (Xu).

Data availability:

All datasets analyzed in this paper are publicly accessible upon request.

Abbreviations Description

HVLTTOT	Hopkins Verbal Learning Test-Revised total learning across trials 1–3
HVLTDELR	Hopkins Verbal Learning Test-Revised delayed free recall
RAVTOT	Rey Auditory Verbal Learning Test total learning across trials 1–5
RAVIR	Rey Auditory Verbal Learning Test total learning immediate recall
RAVDELR	Rey Auditory Verbal Learning Test total learning delayed recall
SYDM	Symbol Digit Modalities (total correct)
STRP1CV	Stroop color-word test-Comalli version color naming trial-time to complete [seconds]
STRP2CV	Stroop color-word test-Comalli version word reading trial-time to complete [seconds]
STRP3CV	Stroop color-word test-Comalli version color-word interference trial-time to complete [seconds]
STRP1GV	Stroop color-word test-Golden version color naming read in 45 seconds

STRP2GV	Stroop color-word test-Golden version word reading read in 45 seconds
STRP3GV	Stroop color-word test-Golden version color-word interference read in 45 seconds
TRLA	Trail Making Test: Part A time to complete [seconds]
TRLB	Trail Making Test: Part B time to complete [seconds]
PASAT	Paced Auditory Serial Addition Task - 50-item (total correct)
GPEGDOM	Grooved Pegboard Test dominant hand-time to complete [seconds]
GPEGNDOM	Grooved Pegboard Test NON-dominant hand-time to complete [seconds]
SEMFLU	Animal fluency (total correct)
VFLU	Phonemic Fluency--FAS (NC08) (total correct)
AFLU	Action Fluency (total correct)

References:

[1]. Sundermann EE, Dastgheyb R, Moore DJ, Buchholz AS, Bondi MW, Ellis RJ, Letendre SL, Heaton RK and Rubin LH, “Identifying and distinguishing cognitive profiles among virally suppressed people with HIV,” Neuropsychology, vol. 38, p. 169–183, 2024. [PubMed: 37971860]

[2]. Rubin LH, Saylor D, Nakigozi G, Nakasujja N, Robertson K, Kisakye A, Batte J, Mayanja R, Anok A, Lofgren SM, Boulware DR, Dastgheyb R, Reynolds SJ, Quinn TC, Gray RH, Wawer MJ and Sacktor N, “Heterogeneity in neurocognitive change trajectories among people with HIV starting antiretroviral therapy in Rakai, Uganda,” Journal of Neurovirology, vol. 25, p. 800–813, 2019. [PubMed: 31218522]

[3]. Rubin LH, Sundermann EE, Dastgheyb R, Buchholz AS, Pasipanodya E, Heaton RK, Grant I, Ellis R and Moore DJ, “Sex Differences in the Patterns and Predictors of Cognitive Function in HIV,” Frontiers in Neurology, vol. 11, p. 551921, 2020. [PubMed: 33329301]

[4]. Dastgheyb RM, Buchholz AS, Fitzgerald KC, Xu Y, Williams DW, Springer G, Anastos K, Gustafson DR, Spence AB, Adimora AA, Waldrop D, Vance DE, Milam J, Bolivar H, Weber KM, Haughey NJ, Maki PM and Rubin LH, “Patterns and Predictors of Cognitive Function Among Virally Suppressed Women With HIV,” Frontiers in Neurology, vol. 12, p. 604984, 2021. [PubMed: 33679577]

[5]. Paul RH, Cho K, Belden A, Carrico AW, Martin E, Bolzenius J, Luckett P, Cooley SA, Mannarino J, Gilman JM, Miano M and Ances BM, “Cognitive Phenotypes of HIV Defined Using a Novel Data-driven Approach,” Journal of Neuroimmune Pharmacology, vol. 17, p. 515–525, 2022. [PubMed: 34981318]

[6]. Dastgheyb RM, Sacktor N, Franklin D, Letendre S, Marcotte T, Heaton R, Grant I, McArthur JC, Rubin LH and Haughey NJ, “Cognitive Trajectory Phenotypes in Human Immunodeficiency Virus-Infected Patients,” Journal of Acquired Immune Deficiency Syndromes (1999), vol. 82, p. 61–70, 2019. [PubMed: 31107302]

[7]. Vonk JMJ, Gross AL, Zammit AR, Bertola L, Avila JF, Jutten RJ, Gaynor LS, Suemoto CK, Kobayashi LC, O’Connell ME, Elugbadebo O, Amofa PA, Staffaroni AM, Rentería MA, Turney IC, Jones RN, Manly JJ, Lee J and Zahodne LB, “Cross-national harmonization of cognitive measures across HRS HCAP (USA) and LASI-DAD (India),” PLoS ONE, vol. 17, p. e0264166, 2022.

[8]. Griffith L, van den Heuvel ER, Fortier I, Sohel N, Hofer S, Payette H, Wolfson C, Belleville S, Kenny M, Doiron D and Raina P, "Statistical approaches to harmonize data on cognitive measures in systematic reviews are rarely reported.," Journal of clinical epidemiology, vol. 68 2, pp. 154–62, 2015. [PubMed: 25497980]

[9]. Bauer DJ and Hussong AM, "Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models," Psychological Methods, vol. 14, p. 101–125, 2009. [PubMed: 19485624]

[10]. Kolen MJ and Brennan R, Test Equating, Scaling, and Linking: Methods and Practices, Springer, 2004.

[11]. Gross AL, Inouye SK, Rebok GW, Brandt J, Crane PK, Parisi JM, Tommet D, Bandeen-Roche K, Carlson MC and Jones RN, "Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time," Journal of Clinical and Experimental Neuropsychology, vol. 34, p. 758–772, 2012. [PubMed: 22540849]

[12]. Adimora AA, Ramirez C, Benning L, Greenblatt RM, Kempf M-C, Tien PC, Kassaye S, Anastos K, Cohen M, Minkoff H, Wingood G, Ofotokun I, Fischl MA and Gange S, "Cohort Profile: The Women's Interagency HIV Study (WIHS)," International Journal of Epidemiology, vol. 47, p. 393–394i, 2018. [PubMed: 29688497]

[13]. Bacon MC, von Wyl V, Alden C, Sharp G, Robison E, Hessol N, Gange S, Barranday Y, Holman S, Weber K and Young MA, "The Women's Interagency HIV Study: an observational cohort brings clinical sciences to the bench," Clinical and Diagnostic Laboratory Immunology, vol. 12, p. 1013–1019, 2005. [PubMed: 16148165]

[14]. Barkan SE, Melnick SL, Preston-Martin S, Weber K, Kalish LA, Miotti P, Young M, Greenblatt R, Sacks H and Feldman J, "The Women's Interagency HIV Study," Epidemiology, vol. 9, p. 117–125, March 1998. [PubMed: 9504278]

[15]. Becker JT, Kingsley LA, Molsberry S, Reynolds S, Aronow A, Levine AJ, Martin E, Miller EN, Munro CA, Ragin A, Sacktor N and Selnes OA, "Cohort Profile: Recruitment cohorts in the neuropsychological substudy of the Multicenter AIDS Cohort Study," International Journal of Epidemiology, vol. 44, p. 1506–1516, 2015. [PubMed: 24771276]

[16]. Morgello S, Gelman BB, Kozlowski PB, Vinters HV, Masliah E, Cornford M, Cavert W, Marra C, Grant I and Singer EJ, "The National NeuroAIDS Tissue Consortium:a new paradigm in brain banking with an emphasis on infectious disease," Neuropathology and Applied Neurobiology, vol. 27, pp. 326–335, 2001. [PubMed: 11532163]

[17]. Rubin LH, Maki PM, Du Y, Sweeney SE, O'Toole R, Nam H, Lee H, Soule AR, Rowe SP, Lesniak WG, Minn I, Dastgheyb R, Shorer EF, Wugalter KA, Severson J, Wu Y, Hall AW, Mathews WB, Kassiou M, Dannals RF, Kassaye SG, Brown TT, Bakker A, Pomper MG and Coughlin JM, "Imaging the translocator protein 18 kDa within cognitive control and declarative memory circuits in virally-suppressed people with HIV," AIDS, 2024.

[18]. May PE, Heithoff AJ, Wichman CS, Phatak VS, Moore DJ, Heaton RK and Fox HS, "Assessing Cognitive Functioning in People Living With HIV (PLWH): Factor Analytic Results From CHARTER and NNTC Cohorts," JAIDS Journal of Acquired Immune Deficiency Syndromes, vol. 83, p. 251–259, March 2020. [PubMed: 31913991]

[19]. von Davier M and von Davier AA, "A UNIFIED APPROACH TO IRT SCALE LINKING AND SCALE TRANSFORMATIONS," ETS Research Report Series, vol. 2004, pp. i–21, 2004.

[20]. Cuschieri S, "The STROBE guidelines," Saudi Journal of Anaesthesia, vol. 13, p. S31–S34, 2019. [PubMed: 30930717]

[21]. Hu L and Bentler PM, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," Structural Equation Modeling: A Multidisciplinary Journal, vol. 6, pp. 1–55, 1999.

[22]. Schermelleh-Engel K, Moosbrugger H and Müller H, "Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures," Methods of Psychological Research, vol. 8, p. 23–74, 2003.

[23]. Hooper D, Coughlan J and Mullen M, "Structural Equation Modeling: Guidelines for Determining Model Fit," The Electronic Journal of Business Research Methods, vol. 6, November 2007.

[24]. Brunner M, Nagy G and Wilhelm O, “A Tutorial on Hierarchically Structured Constructs,” Journal of Personality, vol. 80, pp. 796–846, 2012. [PubMed: 22091867]

[25]. Revelle William, “psych: Procedures for Psychological, Psychometric, and Personality Research,” Evanston, 2024.

[26]. R Core Team, “R: A Language and Environment for Statistical Computing,” Vienna, 2023.

[27]. Muthén LK and Muthén BO, “Mplus: Statistical Analysis with Latent Variables: User’s Guide,” 2017.

[28]. Van der Willik KD, Licher S, Vinke EJ, Knol MJ, Darweesh SKL, van der Geest JN, Schagen SB, Ikram MK, Luik AI and Ikram MA, “Trajectories of Cognitive and Motor Function Between Ages 45 and 90 Years: A Population-Based Study,” The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, vol. 76, pp. 297–306, 2021. [PubMed: 32750110]

[29]. Nichols EL, Cadar D, Lee J, Jones RN and Gross AL, “Linear linking for related traits (LLRT): A novel method for the harmonization of cognitive domains with no or few common items,” Methods, vol. 204, pp. 179–188, 2022. [PubMed: 34843977]

[30]. Livingston SA, Equating Test Scores (without IRT), 2 ed., Educational Testing Service, 2014.

[31]. Wiberg M and Bränberg K, “Kernel Equating Under the Non-Equivalent Groups With Covariates Design,” Applied Psychological Measurement, vol. 39, p. 349–361, July 2015. [PubMed: 29881012]

[32]. Wallin G and Wiberg M, “Nonequivalent Groups with Covariates Design Using Propensity Scores for Kernel Equating,” in Quantitative Psychology, Cham, 2017.

What is new?

- We developed a refined method for harmonizing cognitive data across several large-scale studies in PWH that used different cognitive batteries with only some overlapping tests, unlike traditional methods that require common tests as linking tests.
- The harmonized scores accurately reflect variations, according to age and education status, while preserving the longitudinal cognitive trajectories of participants.
- Our harmonization methods are essential for addressing future analyses to understand the heterogeneity in cognitive complications in PWH.
- These results and parameters can be used to harmonize new datasets with similar tests.

Highlights

- Developed a refined method for harmonizing longitudinal cognitive data in HIV
- Refined approach uses the structural relationships among cognitive domains
- Harmonized cognitive domain scores strongly correlate with raw test scores
- Harmonized scores show similar patterns by demographics and HIV-status across cohorts
- Methods may enhance clinical utility by promoting generalizability in the study of cognition in HIV
- Approach facilitates the ability to address future analyses to understand the heterogeneity in cognitive complication in HIV

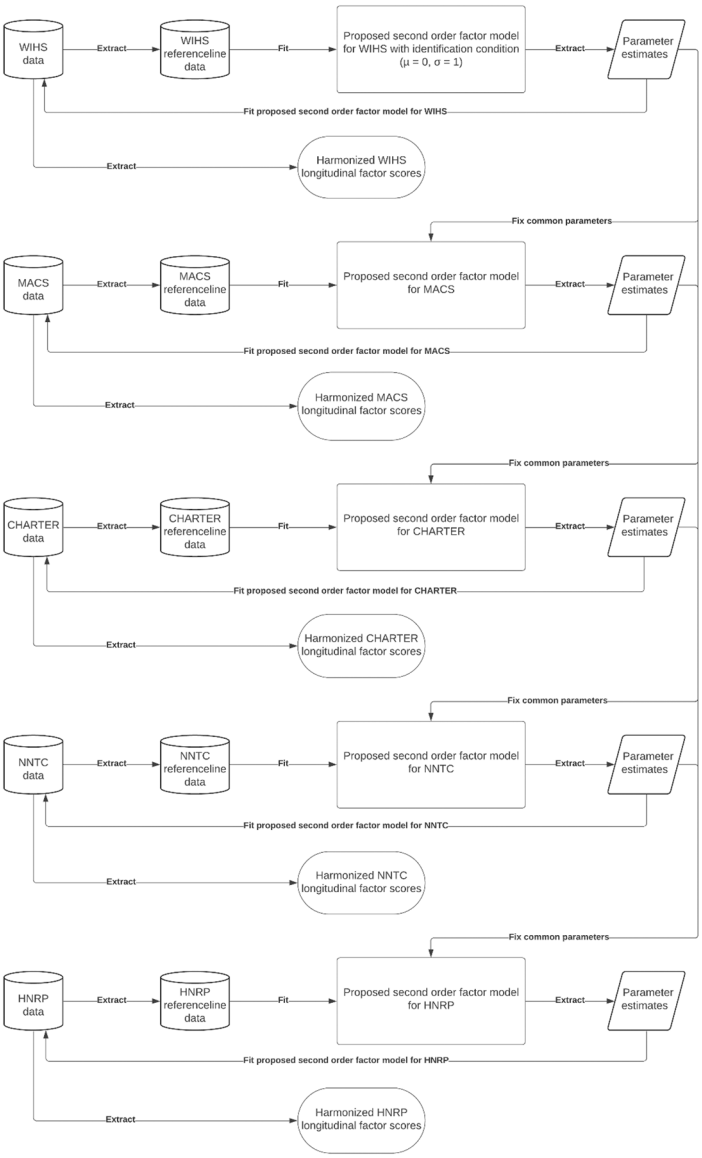


Figure 1.
Harmonization procedure.

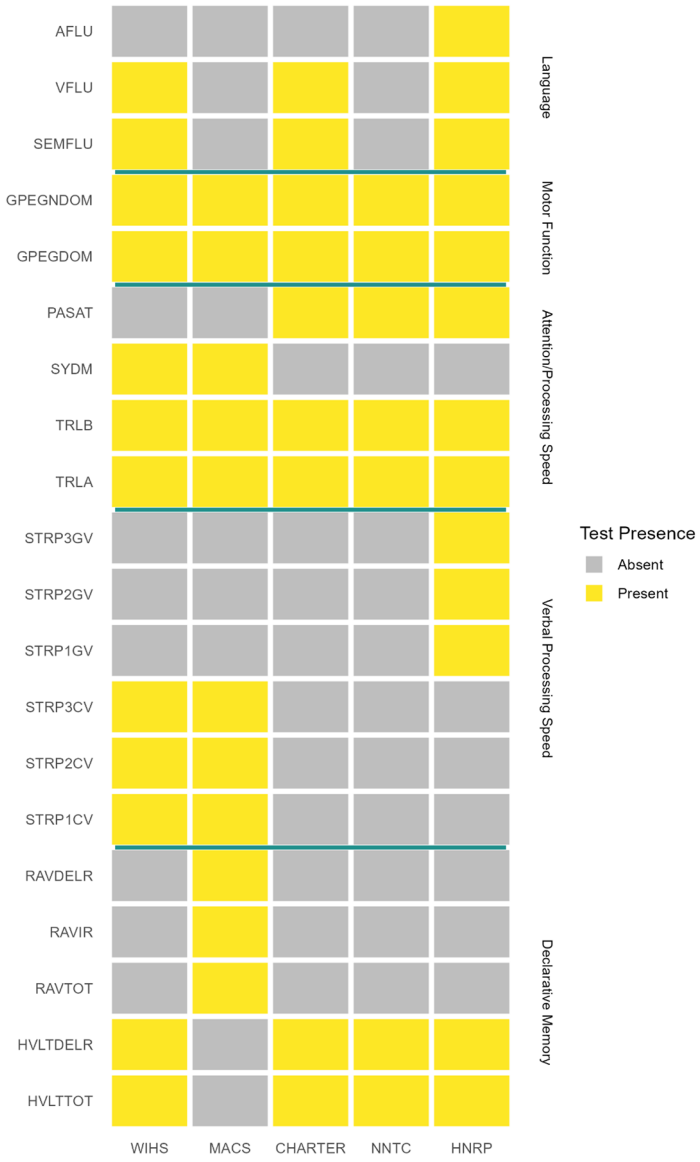


Figure 2.
Cognitive tests included in the data harmonization process.

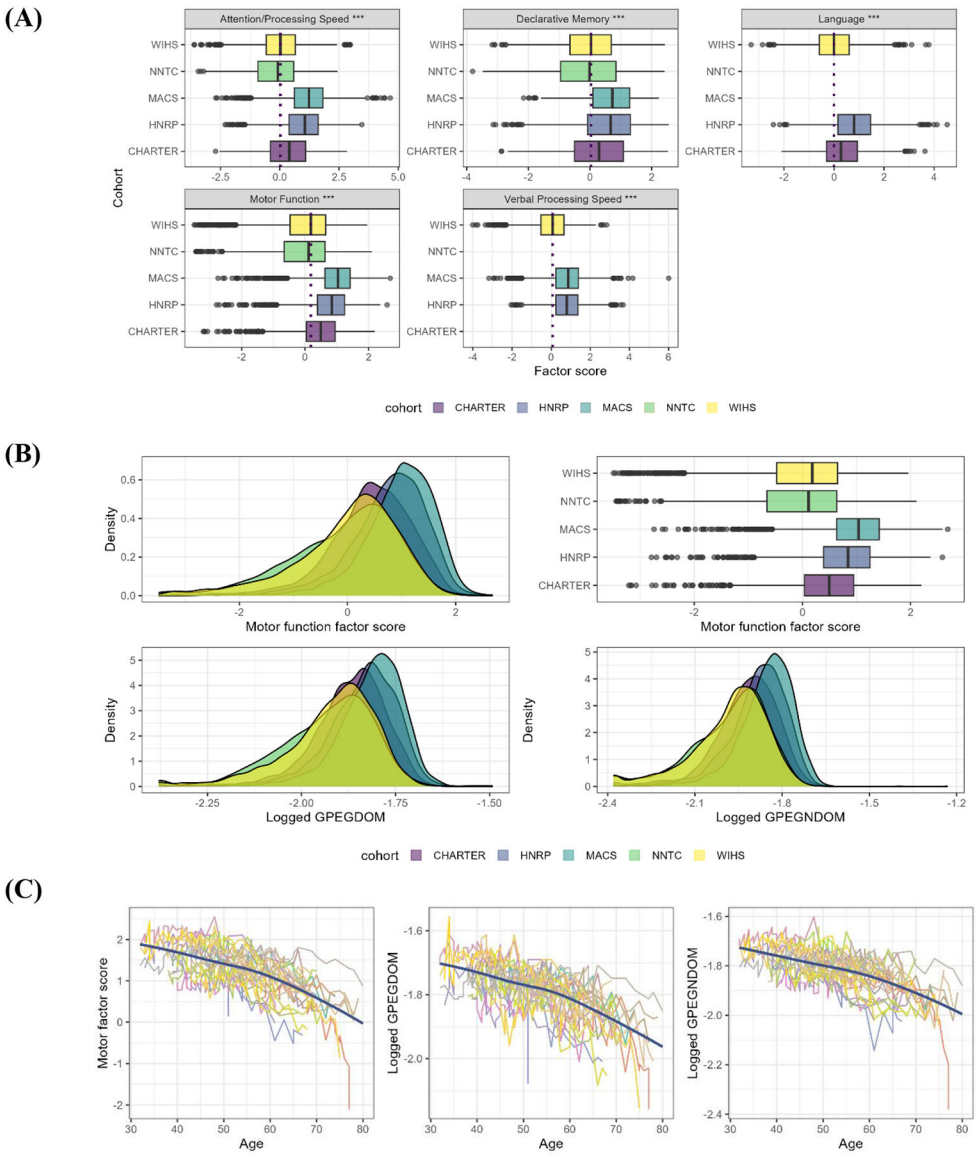


Figure 3.

A) Harmonized factor scores for the five HIV cohort; **B**) harmonized motor function factor scores distribution compared with the raw tests scores distribution; and **C**) longitudinal harmonized motor factor scores for participants with more than 33 visits in MACS. Each colored line represents one participant, the x axis shows the age of the participant at the corresponding visit and the y axis shows harmonized motor factor score or Grooved Pegboard score. The dark blue line is the LOESS line. GPEGDOM=Grooved Pegboard-dominant hand; GPEGNDOM=Grooved Pegboard-nondominant hand ***P<0.01 for ANOVA test.

Table 1.

Demographic characteristics by cohort at the reference line visit (first visit where each participant has complete demographic and cognitive data).

	WIHS[a] (N=2637) n (%)	MACS[b] (N=3635) n (%)	CHARTER[c] (N=1528) n (%)	NNTC[d] (N=1321) n (%)	HNRP[e] (N=2024) n (%)	Overall (N=11145) n (%)	*P*-value	Post-hoc
Median number of visits (IQR)	2 (2)	5 (6)	2 (3)	3 (5)	1 (2)	3 (4)	<0.01	b>d>a>c>e
Proportion of visits completed							<0.01	
1	497 (18.8)	544 (15.0)	738 (48.3)	345 (26.1)	1048 (51.8)	3172 (28.5)		e,c>d>a>b
2	916 (34.7)	547 (15.0)	262 (17.1)	227 (17.2)	355 (17.5)	2307 (20.7)		a>e,d,c,b
3	352 (l3.3)	383 (10.5)	86 (5.6)	144 (10.9)	204 (10.1)	1169 (10.5)		a,d,b,e>c
>3	872 (33.1)	2161 (59.4)	442 (28.9)	605 (45.8)	417 (20.6)	4497 (40.3)		b>d>a>c>e
Mean age (SD)	45.5 (9.22)	41.3 (9.97)	43.0 (8.57)	47.9 (11.1)	44.4 (12.1)	43.9 (10.4)	<0.01	d>a>e>c>b
Male	0 (0)	3635 (100)	1174 (76.8)	673 (50.9)	1594 (78.8)	7076 (63.5)	<0.01	b>e,c>d>a
High school or above	1807 (68.5)	3388 (93.2)	1039 (68.0)	943 (71.4)	1704 (84.2)	8881 (79.7)	<0.01	b>e>d,a,c
Black	1929 (73.2)	904 (24.9)	749 (49.0)	439 (33.2)	303 (15.0)	4324 (38.8)	<0.01	a>c>d>b>e
Hispanic	376 (14.3)	424 (11.7)	190 (12.4)	349 (26.4)	375 (18.5)	1714 (15.4)	<0.01	d>e>a,c,b
People with HIV	1809 (68.6)	2078 (57.2)	1528 (100)	1316 (99.6)	1216 (60.1)	7947 (71.3)	<0.01	c,d>a>e,b

Note. ANOVA was used to compare cohorts on normally distributed continuous factors (nonparametric for non-normally distributed data where median and interquartile range [IQR] are noted) and Chi-square test for categorical factors. Post-hoc ordering is obtained through pairwise two sample t-test or pairwise Fisher's exact test between cohorts and Bonferroni correction.

Table 2.

Cognitive tests administered in at least one of the five harmonization cohorts. Blue squares indicate the test was administered within the cohort.

Test (outcomes available)	WIHS	MACS	CHARTER	NNTC	HNRP
Hopkins Verbal Learning Test-Revised (total learning across trials 1–3, delayed free recall)	•		•	•	•
Rey Auditory Verbal Learning Test (total learning across trials 1–5, immediate & delayed free recall)		•			
Stroop color-word test-Comalli version (color naming, word reading, & color-word interference trials-time to complete [seconds])†	•	•			
Stroop color-word test-Golden version (color naming, word reading, & color-word interference trials-stimuli read in 45 seconds)†					•
Trail Making Test: Part A and B (time to complete [seconds])†	•	•	•	•	•
Grooved Pegboard Test (dominant & non-dominant hand-time to complete [seconds])†	•	•	•	•	•
Symbol Digit Modalities (total correct)	•	•			
WAIS-III Digit Symbol (total correct)			•	•	•
Wisconsin Card Sorting Test – 64 Card Version (total number of perseverative responses)†			•	•	•
Paced Auditory Serial Addition Task - 50-item (total correct)			•	•	•
Controlled Oral Word Association Test (total correct)	•		•	•	•
Animal fluency (total correct)	•		•		•
Action Fluency (total correct)					•

†Note. Outcomes are log transformed and reverse scored so that higher equates to better performance for data harmonization.

Table 3.

Summary for harmonized domain scores across cohorts at the reference line visit (first visit where each participant has complete demographic and cognitive data).

	WIHS[a] (N=2637)	MACS[b] (N=3635)	CHARTER[c] (N=1528)	NNTC[d] (N=1321)	HNRP[e] (N=2024)	*P*-value	Post-hoc
Declarative Memory	0.001 (0.948)	0.67 (0.817)	0.262 (1.082)	−0.09 (1.176)	0.576 (1.001)	<0.01	b>e>c>a,d
Verbal Processing Speed	0.003 (0.931)	0.796 (0.882)			0.803 (0.849)	<0.01	e,b>a
Attention/Processing Speed	0.009 (0.937)	1.191 (0.943)	0.347 (1.006)	−0.174 (1.068)	0.989 (0.905)	<0.01	b>e>c>a>d
Motor function	0.001 (0.947)	0.97 (0.644)	0.43 (0.758)	−0.067 (0.98)	0.752 (0.709)	<0.01	b>e>c>a,d
Language	0.004 (0.889)		0.313 (0.896)		0.817 (0.972)	<0.01	e>c>a

Note. The entries under each cohort represent the mean and standard deviation of each domain factor in that cohort. ANOVA was used to compare cohorts on the domain factors. Post-hoc ordering is obtained through pairwise two sample t-test between cohorts and Bonferroni correction.