University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Kennedy, E;Vadlamani, S;Lindsey, HM;Lei, P-W;Jo-Pugh, M;Thompson, PM;Tate, DF;Hillary, FG;Dennis, EL;Wilde, EA;ENIGMA Clinical Endpoints Working Group,

Title:
Bridging big data in the ENIGMA consortium to combine non-equivalent cognitive measures

Date:
2024-10-16

Citation:
Kennedy, E., Vadlamani, S., Lindsey, H. M., Lei, P. -W., Jo-Pugh, M., Thompson, P. M., Tate, D. F., Hillary, F. G., Dennis, E. L., Wilde, E. A. & ENIGMA Clinical Endpoints Working Group, (2024). Bridging big data in the ENIGMA consortium to combine non-equivalent cognitive measures. Scientific Reports, 14 (1), https://doi.org/10.1038/s41598-024-72968-x.

Persistent Link:
https://hdl.handle.net/11343/353935

License:
CC BY-NC-ND

scientific reports

Check for updates

OPEN

Bridging big data in the ENIGMA consortium to combine non-equivalent cognitive measures

Eamonn Kennedy[1,2,3]✉, Shashank Vadlamani[1], Hannah M. Lindsey[1,3], Pui-Wa Lei[4], Mary Jo-Pugh[1,2], Paul M. Thompson[5,6], David F. Tate[1,3], Frank G. Hillary[7,8,9], Emily L. Dennis[1,3], Elisabeth A. Wilde[1,3] & for the ENIGMA Clinical Endpoints Working Group[118]

Investigators in neuroscience have turned to Big Data to address replication and reliability issues by increasing sample size. These efforts unveil new questions about how to integrate data across distinct sources and instruments. The goal of this study was to link scores across common auditory verbal learning tasks (AVLTs). This international secondary analysis aggregated multisite raw data for AVLTs across 53 studies totaling 10,505 individuals. Using the ComBat-GAM algorithm, we isolated and removed the component of memory scores associated with site effects while preserving instrumental effects. After adjustment, a continuous item response theory model used multiple memory items of varying difficulty to estimate each individual's latent verbal learning ability on a single scale. Equivalent raw scores across AVLTs were then found by linking individuals through the ability scale. Harmonization reduced total cross-site score variance by 37% while preserving meaningful memory effects. Age had the largest impact on scores overall (− 11.4%), while race/ethnicity variable was not significant ($p > 0.05$). The resulting tools were validated on dually administered tests. The conversion tool is available online so researchers and clinicians can convert memory scores across instruments. This work demonstrates that global harmonization initiatives can address reproducibility challenges across the behavioral sciences.

Keywords Harmonization, Verbal learning, Mega analysis, Traumatic brain injury, Item response theory

Data sharing consortia aim to increase the robustness and statistical power of results by aggregating large and diverse samples[1,2]. While analyses of large datasets can provide unparalleled statistical power, data aggregation without robust harmonization can mask and even introduce flaws and biases[3]. This is a critical consideration in the behavioral sciences, where multi-site collaboration often requires the synthesis of non-identical cognitive measures[4–6]. For example, verbal memory/recall is a core cognitive function, and deficits in learning and memory are one of the most common and widely assessed patient complaints[7]. However, a wide variety of auditory verbal learning tasks (AVLTs) exist that can be administered to assess verbal memory and recall, and these differ across a range of qualitative and quantitative features[7–9]. Such differences in assessment instruments can contribute to inconsistencies in the measurement of neurocognitive performance[10]. Therefore, what is needed is a means to accurately convert scores across common AVLTs. This paper specifically focused on auditory verbal learning tests (AVLTs) as an example of how harmonisation procedures previously applied to other assessments[4]. can be applied to harmonize instruments commonly used to assess cognitive effects following TBI.

Methods to accurately relate scores across AVLTs could facilitate highly powered studies of verbal memory and recall, offering opportunities for new clinical insights[11,12]. However, data from single sites are typically

[1]Department of Neurology, University of Utah School of Medicine, Salt Lake City, UT, USA. [2]Division of Epidemiology, University of Utah, Salt Lake City, UT, USA. [3]George E. Wahlen Veterans Affairs Medical Center, Salt Lake City, UT, USA. [4]Department of Educational Psychology, Counseling, and Special Education, Pennsylvania State University, University Park, PA, USA. [5]Imaging Genetics Center, Stevens Neuroimaging & Informatics Institute, Keck School of Medicine of USC, Marina del Rey, CA, USA. [6]Departments of Neurology, Pediatrics, Psychiatry, Radiology, Engineering, and Ophthalmology, USC, Los Angeles, CA, USA. [7]Department of Psychology, Penn State University, State College, PA, USA. [8]Department of Neurology, Hershey Medical Center, State College, PA, USA. [9]Social Life and Engineering Science Imaging Center, Penn State University, State College, PA, USA. [118]A comprehensive list of consortium members appears at the end of the paper. ✉email: eamonn.kennedy@utah.edu

biased by the specific attributes, demographics, and inclusion criteria of the study which can increase variance/error and confound reproducibility[13–15]. To address these limitations, emerging data harmonization approaches offer new ways to perform data transformations that remove unwanted influences in aggregated data, such as site-specific differences in test administration, while preserving meaningful effects. Data harmonization of large and heterogeneous AVLT data sources represents an appropriate framework for the development of cross-AVLT score conversion tools, but such efforts come with analytical and organizational challenges[1,4,16,17].

In particular, the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Brain Injury working group brings together researchers from around the world to study brain structure, function, and cognitive endpoints after brain injury by aggregating large sample studies from multiple studies. However, studies that aim to evaluate subtle differences in cognitive endpoints of brain injury must overcome three primary challenges (1) A large, international sample of data is required including both controls and brain injured individuals, (2) Multisite raw data aggregation requires methods that can isolate and remove unwanted site effects, while explicitly preserving meaningful relationships, (3) Appropriate psychometric methods are needed to measure and account for different instrumental effects across multiple item scales.

To overcome these challenges, we report a retrospective multisite (n = 53 datasets) mega study analysis of three Common AVLTs: the California Verbal Learning Test (CVLT[7]; the Rey Auditory Verbal Learning Test (RAVLT[18]; and the Hopkins Verbal Learning Test-Revised (HVLT[19]; drawing from international healthy and brain-injured populations across 13 countries and 8 languages. In contrast to meta-analyses which combine summary statistics from several sites, we conducted a mega-analysis that centralizes and pools individual raw data from many sites. This allows for a richer range of experimental designs which can consider subtle single item differences in detail[1,2].

Our primary hypothesis was that conversion performance would be significantly improved by a mega-analytic pipeline combining harmonization and item response theory (IRT) models. IRT is an appropriate method for this purpose because it can make use of multiple memory items of varying properties and difficulty in order to place all individuals on the same ability scale, regardless of the memory instrument used for assessment. Similarly, batch harmonization algorithms are appropriate when there may be spurious or nonbiological effects attributed to a large number of underlying sites that must be isolated and removed. The goal of this study was to establish crosswalks between common memory measures, and address long standing data compatibility issues for AVLTs through the dissemination of freely available instrument conversion tools: enigma-tools.shinyapps.io/verbal-learning-calculator/

Methods

Data sources and inclusion criteria

A range of international studies of head injury and comparator groups and controls for a variety of conditions were included. Comprehensive details and references for these studies are provided in the supplement, alongside exhaustive study-level definitions of what constituted brain injury, controls, groups, and inclusion/exclusion criteria (see Supplementary Table S1 and S2)[20–24]. This secondary multisite (N = 53 datasets) mega-analysis focused on three AVLTs: the CVLT[7], HVLT[19], and RAVLT[18]. To mitigate balance issues, we included only comparator controls and groups with TBI. As described in prior work[4], we aggregated data contributed by collaborators in the Psychiatric Genomics Consortium (PGC), the Enhancing NeuroImaging Genetics through Meta-Analysis Consortium (ENIGMA) working groups[25], the ENIGMA Brain Injury working group[15], and the Long-term Impact of Military-relevant Brain Injury Consortium—Chronic Effects of Neurotrauma Consortium, LIMBIC-CENC[17]. The University of Utah provided overall Institutional Review Board (IRB) study approval and each contributed study was approved by the IRBs of their respective institutions. Each contributed study was conducted in accordance with the Declaration of Helsinki, including obtaining informed consent from each participant.

To limit sources of variability, we excluded anyone with a known clinically diagnosed mental health or neurological condition other than traumatic brain injury (TBI). Consistent with standard AVLT administration practices, we included only participants aged 16 years or over. In the case of longitudinal or serial measurement designs, only the first measurement of AVLTs per person were included; repeated measurements were dropped.

Verbal learning task contents and scoring

Table 1 provides a summary overview of the key features of the AVLTs assessed. AVLT scores on each trial denote the number of correct words that are recalled. The maximum score reflects the number of memory items per list. The sum of the total words recalled across all immediate free recall (learning) trials is the immediate free recall summary score (Sum of Learning Trials). These raw scores are often subsequently normed so that the performance of the individual can be contextualized relative to a population of interest. However, in this work we exclusively assess raw scores, and not t-scores or normative scores. We focused on raw scores because normative values are occasionally updated over time and are based upon instrumentally-distinct normative samples.

The California verbal learning test

The CVLT[7] refers to a family of instruments that assess verbal learning and memory deficits. The CVLT has been revised twice, and three iterations exist (CVLT-I, CVLT-II, and CVLT-3). Additionally, the CVLT comes in standard, short, and alternate forms. In this work, we estimated crosswalks for the more recent CVLT-II and the CVLT-3. While the CVLT-3 is nominally a revision of the CVLT-II, in practice the target words, their order, and their number are the same for both the CVLT-3 and CVLT-II. Thus, we refer to both CVLT-3 and CVLT-II standard and alternate forms together as 'CVLT'. Table 1 provides a numerical overview of the key features of the CVLT. The CVLT uses M = 16-word list lengths, which are drawn from 4 semantic categories, and 5 consecutive

	CVLT	RAVLT	HVLT
List lengths (words)	16	15	12
Semantic categories	4	–	3
Immediate free recall trials	5	5	3
Recognition trial, total words	48	50	24
Long delay duration (minutes)	20–30	20	20–25
Distractor list (List B)	Yes	Yes	No
Cued recall trials	Yes	No	No
Short Delay Trial	Yes	Yes	No

Table 1. Summary of the key features of the three AVLTs.

learning trials. The CVLT is a comprehensive test that includes a distractor list, cued and free recall assessments, short and long delay trials, and a recognition trial with 48 words.

The Hopkins verbal learning test–revised

The HVLT-R[19] is a relatively short measure of verbal learning and memory deficits. The HVLT exists in two primary forms (original and revised) denoted together as 'HVLT'. Table 1 outlines the key features of the HVLT. The HVLT does not use a distractor list for immediate recall and does not assess cued or a short delay recall performance. The HVLT uses M = 12-word list lengths, which are drawn from 3 semantic categories, and uses a small (N = 24) total pool of words for scoring. The HVLT has three consecutive learning trials.

The rey auditory verbal learning test

The RAVLT[18] is a measure of verbal learning and memory deficits. Table 1 provides an overview of the key features of the RAVLT. The RAVLT draws from random, semantically unrelated words, and employs a M = 15-word list length, a distractor list, as well as a large (N = 50) total pool of words for scoring recognition hits. Alternate forms also exist for the RAVLT.

Covariates

Language, country of origin, age at testing, sex/gender, race/ethnicity, site/study, military/civilian status, TBI history, and education level were included and adjusted for in this study. The exclusion criteria were used to rule out the presence of any other clinically relevant variables, including epilepsy, dementia, and mild or early onset cognitive impairment[26]. While some of the studies recorded gender, others recorded biological sex, and these were aggregated into a single variable. Ethnicity was binarized to Hispanic/Latino, or Not Hispanic/Latino. Perspectives on race/ethnicity differ widely according to cultural context[27], and we elected to use broad categories of Black, White, Asian, and Other. Covariate coefficients per AVLT model were converted to percentages, averaged, and then applied back to adjust the full cohort. This means each covariate had the same effect on scores regardless of the instrument used.

Statistical analysis

Analysis was performed in Python 3 and in R. Kruskal–Wallis *H* tests (omnibus) were used to test for overall significance across groups. Where normality was confirmed, *t* tests were used for post-hoc pairwise comparisons with additional correction for multiple comparisons. Overall missing values were low (< 5%) and any missing data points were imputed with nearest neighbor imputation. After data cleaning and imputation, Empirical Bayes harmonization using the ComBat-GAM algorithm[13] was used to remove unwanted site effects while preserving instrumental effects for further analysis. The ComBat-Gam algorithm is a recent, popular approach for removing complex batch effects while preserving important other data properties.

Correcting for site effects

Modeling was conducted in three stages: first, the overall dataset was divided into three subsets, one per AVLT instrument. The ComBat-GAM algorithm[13] was applied to each of these three subsets separately to remove site effects within each AVLT. ComBat-GAM explictly preserves complex covariate effects while isolating and removing site effects[13] and the effectiveness of ComBat in reducing site and batch effects is independent of data type. ComBat-GAM was selected as the harmonisation method because it can preserve nonlinear covariate effects such as age effects on cognition through generalized additive modeling. After site correction, covariate adjustment was performed as follows: the overall dataset was again divided into three distinct subsets by instrument, and ordinary least squares (OLS) linear models were used to estimate and remove covariates in AVLT linear models.

Sampling structure

Cross-validation is constrained for use in harmonisation studies where the data are altered by a procedure using information from other samples that are not part of each validation set. This raises concern for data leakage, where information about the harmonisation or site effect parameters might influence the validation of the crosswalk after harmonisation. To ensure robust validation, we elected to have an entirely separate test set of

scores for individuals who were dually administered two AVLTs. If the crosswalk model (fit between measure scores) agreed with the empirical data, then this would provide evidence of crosswalk accuracy.

Conversion process

Our goal was to obtain equivalent scores across tests. If two people have the same underlying verbal learning ability, then on average they will obtain equivalent (although not necessarily equal) scores on two tests of the same construct, regardless of their difficulty. Therefore, what is needed is to place individuals on a single construct ability scale. We estimated relative item parameters (including difficulties) for all but false positive items from the available data. We then used maximum likelihood estimation to calculate ability scores per person using both their raw item scores and the estimated item parameters (including difficulties). The estimated item difficulties served as weights for the (log) odds of observing a proportion correct of scores, with more discriminating and/or more difficult items weighted more heavily. Weighted item scores were then summed to estimate individual abilities. With all individuals placed on the same ability scale, item scores were equatable.

Item response theory

A Continuous Response Model (CRM) in the IRT family[28–31] was used to estimate each subject's verbal learning ability because item score ranges were large ($>=12$) and different across tests. CRM is an extension of the Graded Response Model (Samejima, 1969) for continuous response formats. Graded response formats with large number of score points (e.g., $>=9$) are often considered continuous and CRM is appropriate in this condition[32]. To calibrate parameters, we used Shojima's[29] simplified expectation maximization (EM) method by assuming non-informative priors for item parameters as implemented in the EstCRM (Continuous Response Model) R package[28]. After all data adjustments, samples taking different AVLTs were assumed to be randomly equivalent (see Limitations), such that verbal learning ability estimates were placed on the same scale using a 'randomly equivalent groups' linking design. The relative difficulty of all items across tests was taken into account to create a single ability scale, and tables of equivalent AVLT scores were linked through the ability scale.

Anchor items and ability measures

Anchor items similar in format and nature were identified for each of the three required crosswalks (1. RAVLT ↔ CVLT; 2. RAVLT ↔ HVLT; and 3. CVLT ↔ HVLT). After expert consensus and trials of different anchor combinations, we elected to use immediate free recall learning trials, short delay, and long delay free recall as anchor items, where available. Short delay was used as an anchor item between CVLT and RAVLT only (short delay is not assessed in HVLT). False positive measures were not recorded consistently across sites and were not used. Recognition hits showed inconsistent behavior and were excluded from conversions (see Limitations). Since all site effects and measured covariate effects had been removed prior to IRT analysis, we assumed scores were randomly equivalent across measures and ability scores did not require further scaling.

Results

Data summary

An overview of the key features differentiating the CVLT, RAVLT, and HVLT assessments are provided in Table 1. Supplementary Table S1 shows summary characteristics itemized for each of the 53 aggregated datasets after applying exclusion criteria. Overall, the sample size was N = 10,505 (31.8% female) which included both controls and TBI groups. The median age was 42 years with an interquartile range of 30–55 years. Different studies showed significant differences in total sum of trial scores (Fig. 1), and site-related variation in scores was reduced by harmonization. Table 2 shows the summary statistics of the full cohort after aggregation. Each instrument was represented by > 1000 subjects across > 10 studies, indicating good representation of AVLTs. Significant differences in demographic characteristics were evident across measures, indicating that covariate adjustment was required.

Harmonization

The ComBat-GAM algorithm was implemented to correct for site-specific variations such as differences in inclusion/exclusion criteria (Supplementary Table S2) while preserving real covariate effects. Figure 2 shows a comparison of single site mean scores, before (gray dots) and after (colored dots) site harmonization. A line is drawn to connect each site from its pre- to post-harmonized value. Gray distributions portray the variation in mean scores across sites; colored distributions portray site mean score distributions after harmonization (CVLT: Blue, RAVLT: Orange, HVLT: Red). The unadjusted distributions of scores (gray areas) exhibit much higher variance than their post-harmonized equivalents, and overall harmonization reduced total variance by 37% across all items. As covariate effects were preserved, variations owing to unwanted site effect were reduced. In a secondary analysis of latent dimensions, a principal component analysis (PCA) reduction of all verbal learning memory items identified a 3.8% mean reduction in interquartile spread of latent factors after harmonisation across all measures (Supplemental Fig. S1). Therefore, the majority of the latent factor was associated with verbal learning ability, and largely preserved by harmonization.

Covariate adjustment

Figure 3a shows boxplots of unadjusted sum of learning trial scores as percentages stratified by group (TBI vs. control) and sex/gender. TBI history and male sex/gender were both associated with lower sum scores across all three tests. Age-related declines were well-fit by quadratics; years of education were well-fit by a straight line (Fig. 3b,c). The effects of both age and education were significant and consistent across all tests. Linear models within each measure were used to assess covariates (Table 3) after language and country of origin effects were removed separately prior to harmonization.

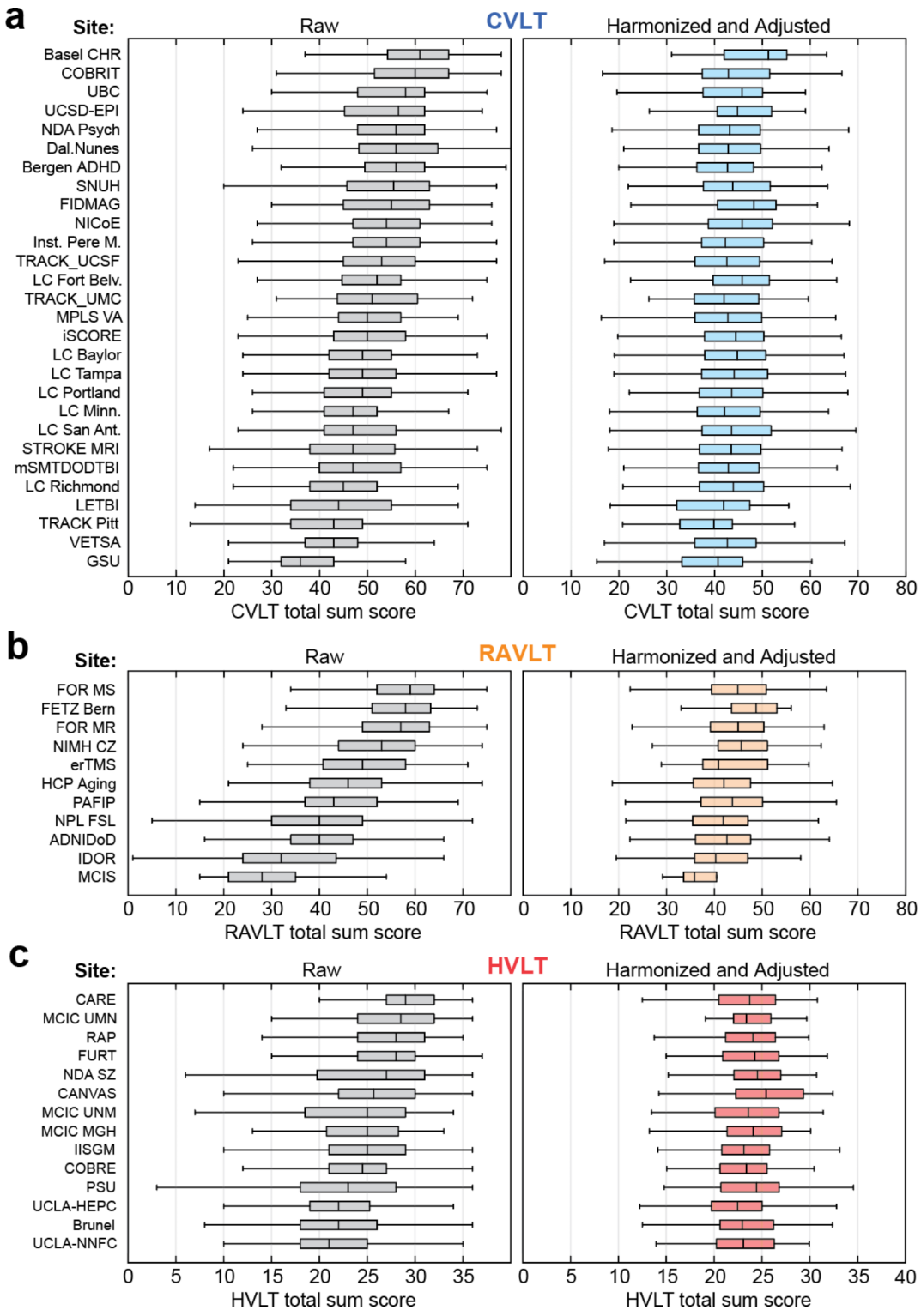


Fig. 1. Comparing multisite data of total of trials scores before and after ComBat harmonization and adjustment for (**a**) CVLT, (**b**) RAVLT, and (**c**) HVLT. Results are sorted by median score per study. Variation in site medians were reduced after harmonization. Full details for all sites are available in Supplementary Fig. S1.

Score conversion

A continuous IRT model (Fig. 4) was used to estimate the latent trait of all individuals while accounting for different item difficulties and discriminations across multiple test items. Third degree polynomials estimated the relationship between observed scores and ability scores for each. Cubic polynomial fits of ability vs. score are shown in Fig. 4a for immediate, short, and long delay items. Horizontal lines of equivalent ability connect

	CVLT	RAVLT	HVLT	*p*
Sample size, N	6634	2728	1143	
# Datasets	28	11	14	
Age: 16–35	34.70%	32.70%	48.20%	< 0.001
35–65	61.20%	40.40%	41.90%	< 0.001
65 +	4.10%	26.90%	9.90%	< 0.001
Gender: Male	76.50%	48.60%	66.10%	< 0.001
Female	23.50%	51.40%	33.90%	< 0.001
Region: Americas	90.70%	44.50%	69.00%	< 0.001
Europe	8.30%	55.40%	27.60%	< 0.001
Asia Pacific	1.10%	0.10%	3.40%	< 0.001
Language: English	90.70%	37.60%	82.20%	< 0.001
Spanish	1.90%	7.40%	6.40%	< 0.001
German	0.20%	35.3%	11.50%	< 0.001
Italian	0.00%	8.80%	0.00%	< 0.001
Other	7.20%	10.90%	0.00%	< 0.001
Education: Secondary	20.80%	28.30%	15.20%	< 0.001
Some College	14.80%	14.00%	19.70%	< 0.001
Postgraduate	7.20%	21.40%	5.50%	< 0.001
Clinical Population: TBI	64.70%	11.90%	21.90%	< 0.001
Controls	35.30%	88.10%	78.10%	< 0.001
Mean Raw Scores: Trial 1	4.9 ± 1.9	5.0 ± 1.9	6.1 ± 1.7	< 0.001
Total Learning	43.2 ± 9.7	42.9 ± 8.6	23.2 ± 4.2	< 0.001
Long Delay	9.6 ± 3.1	8.7 ± 3.0	7.9 ± 2.2	< 0.001

Table 2. Descriptive characteristics of the total cohort by instrument.

equivalent scores. Items with longer delay show larger differences in difficulty than shorter delay items. A secondary sensitivity analysis conducted using only the control population and no individuals with TBI resulted in similar conversions.

Validation with dually administered tests

We validated the derived conversions on held-out data not used in other analyses. Validation was conducted by comparing the conversion estimates to real data where two verbal learning tests were administered to the same set of individuals (Fig. 4b; n = 36). How well conversion lines fit the dually administered test scores is a measure of conversion accuracy. Although this sample size was small compared to the total aggregated data, it still independently suggests agreement between the derived conversion models and dually administered test scores (Fig. 4b, blue shaded area). These data are fitted against the IRT-derived conversion scores (black dotted line) for RAVLT to CVLT. The line falls within the 95% confidence bound for the dually administered tests, indicating agreement. Compared to the same conversion model constructed using unadjusted data, the harmonized conversion exhibited a 9.5% lower root mean squared error against the held-out data, indicating that harmonization moderately improved conversion. Figure 4b indicated plausible, modest model agreement within the constraints of the sample size. As a further validation, cross-item correlations for each AVLT were compared before and after harmonisation. As shown in supplemental Fig. S2, harmonisation had only a small effect on the item cross-correlations (− 0.016 reduction in average cross-correlation). Table 3. Blocked linear regressions predicting sum of learning scores per instrument. The average percentage effect across all AVLTs are shown, indicating that Age > 65 had the largest impact on scores overall (− 11.4%). * indicates significance after correction for multiple comparisons.

Ability scale properties

The properties of the derived verbal learning ability scale were evaluated as a function of sex and TBI status for all participants (Fig. S2). Consistent with prior findings for the total of trials scores (Fig. 3a), females exhibited significantly higher verbal learning ability than males across all instruments. Comparing TBI and control groups, history of TBI was associated with significant declines in ability, ranging from 0.12 to 0.79 standard deviations across groups. In Fig. 5, the distribution of unadjusted ability scores is shown for each site ranked by ability and color-coded by median age per site. Unadjusted median verbal learning ability varied across sites, and these differences were strongly associated with the median age per site, and other covariates.

Application

Details for converting scores using the online tool available at enigma-tools.shinyapps.io/verbal-learning-calculator/are provided as Supplementary Note 1: Procedure for data conversion.

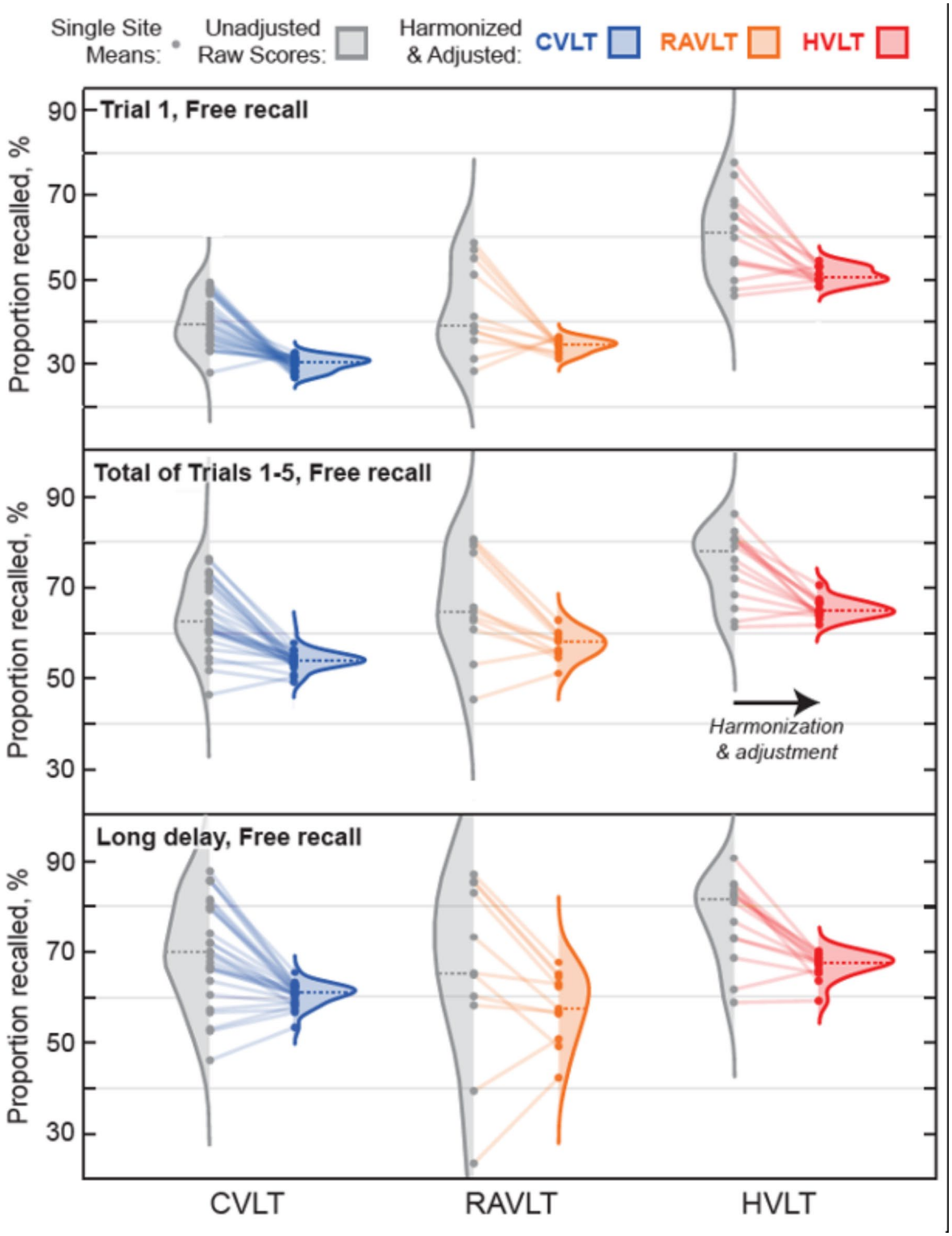


Fig. 2. Comparing proportions of memory items recalled before and after harmonization. Mean scores for each site (dots) are shown broken out by instrument (color) and item (Top: Trial 1 immediate free recall, Middle: Total sum of all Trials, Bottom: Long-delay free recall scores).

Discussion

There have never been more studies published annually in the history of the neurosciences[33] This intensive rate of research offers unparalleled opportunity for data combination and nuanced examination of cognitive and behavioral changes associated with neurological diagnoses. However, "high volume science" lacks coordination between studies, which poses critical challenges for the integration of findings and data harmonization. For

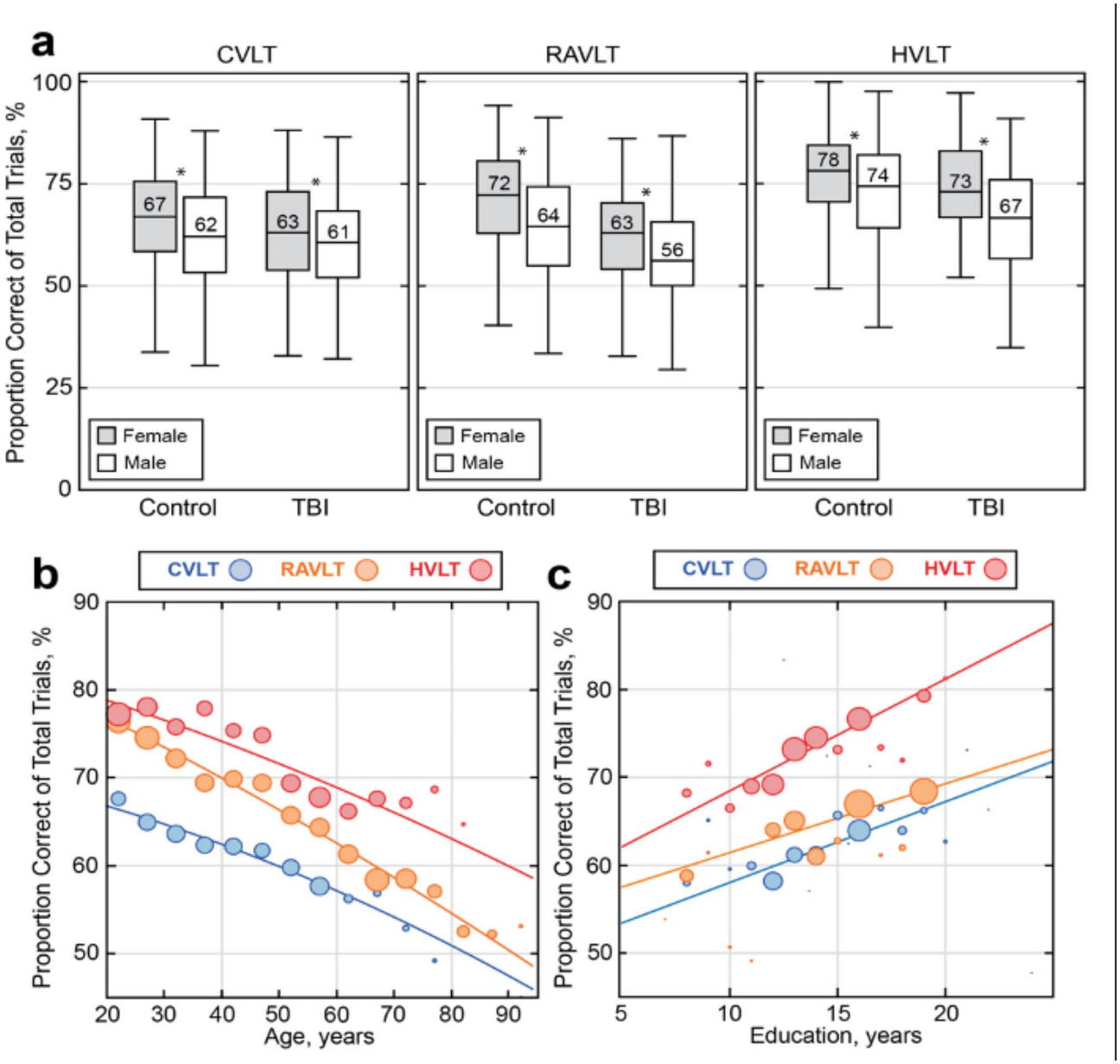


Fig. 3. Visualizing covariate effects on covariate unadjusted, harmonized scores. (**a**) Boxplots of scores stratified by group (TBI vs. control) and sex/gender indicated that males and those with history of TBI had significantly lower scores on average. Age-related declines (**b**) and the beneficial effects of education (**c**) on scores were consistent across all AVLTs.

example, AVLTs are the most common method for learning and memory assessment, but they were independently developed and without explicit quantitative reference to pre-existing instrumentation. Over the last 70 years, this has led to a scenario where clinicians and researchers routinely use distinct AVLTs with incomparable results[7–9]. This is not only a technical inconvenience but is problematic for the interpretation and reproducibility of results and findings.

Constructing reliable standards for converting scores across common AVLTs is challenging, because conversions should be made independent of factors such as language, study group, and instrumental details. For example, given more words to recall, it is more likely that more words will be recalled. Naim et al. found the average number of memory items recalled (R) scales with the root of M items presented[34], and there are other subtle differences between seemingly similar assessments. Large-sample mega-analysis and harmonization present a promising solution to address these concerns and examine interesting clinical features.

Beyond conversions, comparing the difficulty of tests on the same ability scale may assist with the selection of AVLT across different research and clinical contexts. For example, the HVLT was the easiest test overall, while the CVLT was the most challenging test. The HVLT may be most appropriate for the assessment of individuals who are at risk for significant impairment. Conversely, the CVLT has sufficient dynamic range to discriminate

	CVLT	RAVLT	HVLT		
Sample size	6634	2728	1143		
Explained variance	10%	30%	22%	% effect	Significant
Max score of T1-TN	80	75	36	All	All tests
Age group (Ref: 16–35 years)					
35–65	− 2.7*	− 6.5*	− 1.0*	− 4.90%	Yes
65+	− 7.8*	− 12.2*	− 2.9*	− 11.40%	Yes
Sex/gender (Ref: Male)					
Female	3.1*	3.7*	1.6*	4.40%	Yes
Race/Ethnicity (Ref: White)					
Multi	− 1.3	− 3	− 0.6	− 2.40%	No
Black	− 2.7*	− 0.8	− 2.8*	− 4.10%	No
Asian	1.1	0.5	− 1.9*	− 1.10%	No
Hispanic	0.5	− 1.1	− 0.1	− 0.40%	No
TBI (Ref: No TBI)	− 1.3*	− 1.6*	− 1.6*	− 2.80%	Yes
Education (per year)	0.6*	0.6*	0.4*	0.90%	Yes

Table 3. Blocked linear regressions predicting sum of raw learning scores per instrument. The average percentage effect across all AVLTs are shown, indicating that Age > 65 had the largest impact on scores overall (− 11.4%). *Indicates significance after correction for multiple comparisons.

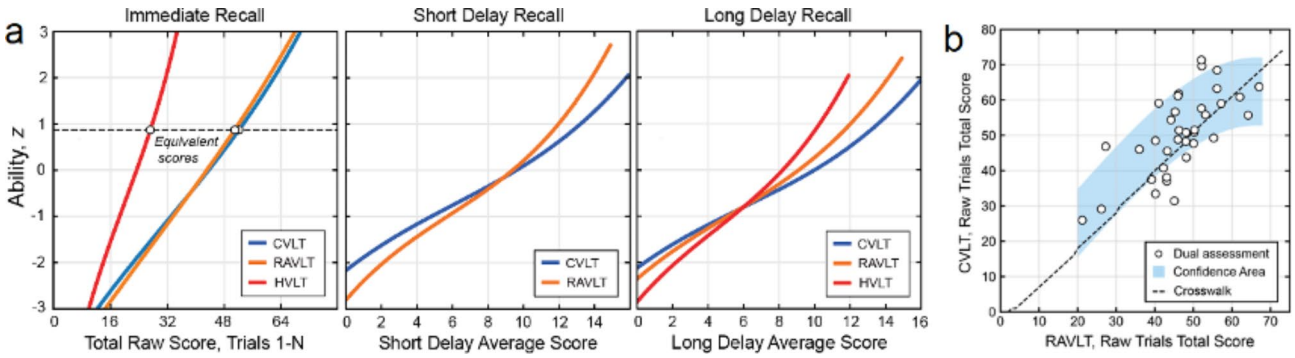


Fig. 4. Visualizing and Validating Conversions. (**a**) Average scores as a function of individual ability are shown approximated as cubic polynomial fits for immediate, short, and long delay trials. Scores shown are not normed or T-scored. Horizontal lines of equivalent ability connect equivalent scores across tests, which facilitates the construction of crosswalks. (**b**) Scatter plot and fit to the sum of learning Trial scores for a subset of cases who were administered both the CVLT and RAVLT (n = 36). The confidence area of the dually assessed data is shown in blue and agrees with the derived crosswalk for CVLT- > RAVLT (n = 9362, black dotted line).

within high ability groups, while the RAVLT may be well suited for studies involving a wide range of abilities. However, these are relatively coarse recommendations which may only be suitable in specific scenarios[35].

In the process of converting across AVLTs, site effects such as different settings, inclusions, and procedures were found to have an appreciable impact on verbal learning scores. This may also be attributable to underlying differences in inclusion/exclusion criteria across studies (Table S2). However, a detailed list of all the ways our sources differed was not necessary to remove these effects in aggregate with a harmonization algorithm. We confirmed our primary hypothesis that conversion error would be reduced by implementing a mega-analytic pipeline combining harmonization and IRT. In time, these conversions may be found to be suitable for clinical utilization at the individual level, although verifying this will require further independent scrutiny.

Appropriate harmonization transforms data in ways that preserve its core relationships. For this study, these relationships include the associations between scores and ability, between scores and covariates (e.g., age-related memory decline), and the measurement of the underlying cognitive construct. After harmonization, the higher scores associated with younger age, female sex/gender, more education, and controls persisted for all AVLTs, despite a large drop in cross-site variance, indicating unwanted effects were removed, while important covariate effects were preserved. Interestingly, despite the large sample size, no Race/Ethnicity variable was consistently associated with higher or lower scores.

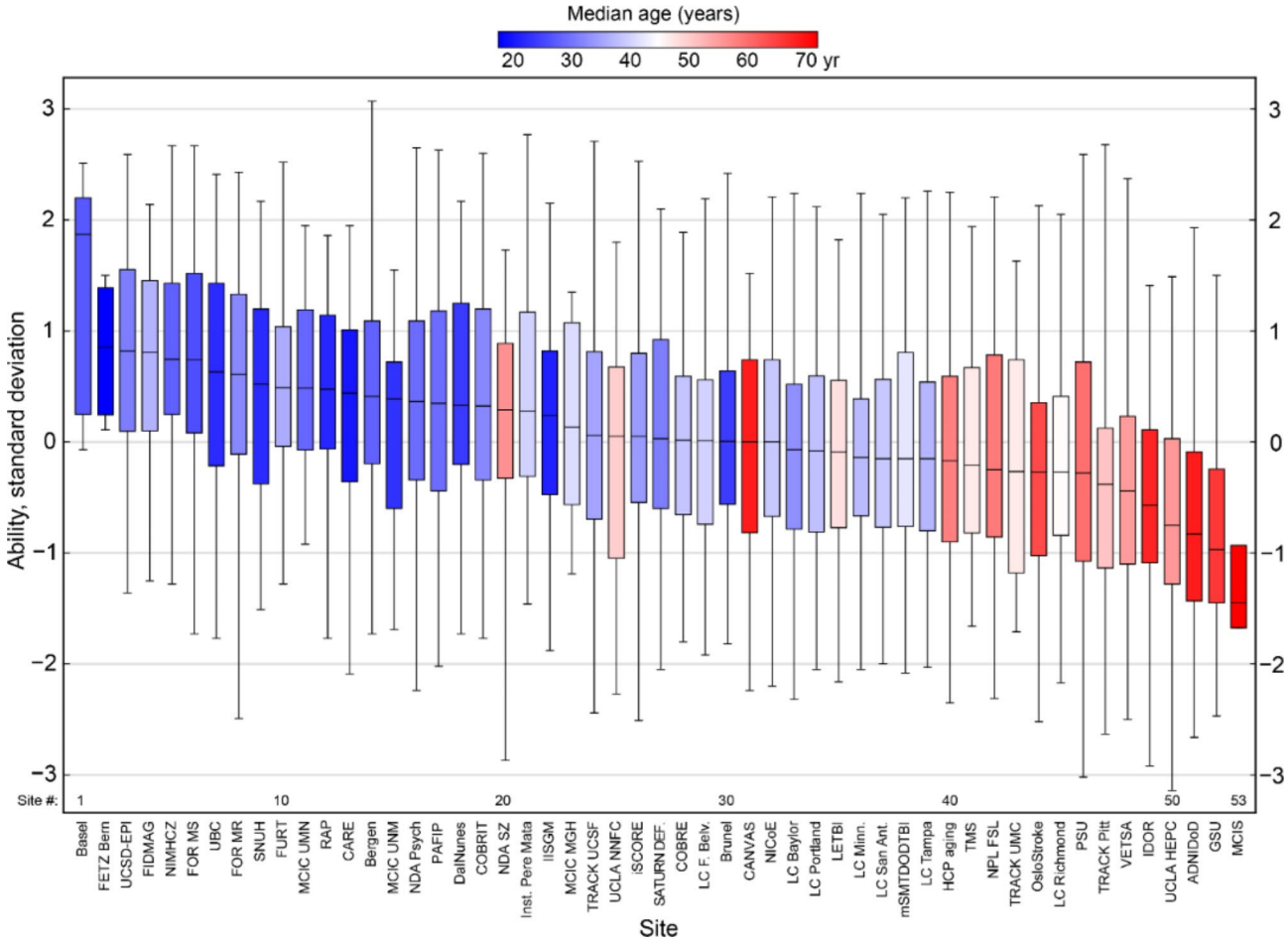


Fig. 5. The distribution of unadjusted ability scores are shown for each site, ranked by ability and color-coded by median age per site.

Strengths and limitations

Strengths of this study include a comprehensive dataset of more than ten thousand participants drawn from 53 international datasets that recorded performance on verbal learning tasks. This work suggests the specific choice of AVLT has a pronounced effect; averaged across items, the CVLT was the most challenging test, although it was similar to the RAVLT in difficulty, while the HVLT was the least difficult, as expected due to its lower complexity[36]. Free conversion tables and tools can assist clinicians to track and compare patient scores against large reference groups, regardless of differences in AVLT administration practices. For example, the derived crosswalks operate independent of whether adjustment has or has not been implemented in the scores input into the online calculator. More broadly, this work demonstrates that data harmonization of large data sharing initiatives can offer new tools to address long standing data challenges.

Our study has several limitations. This study considered only a limited binary interpretation of lifetime history of TBI, which exists along a spectrum of severity and has distinct phenotypes ([37]. However, a secondary sensitivity analysis conducted using only the control population and no individuals with TBI resulted in similar conversions. This study was primarily from English-speaking and western hemisphere countries. Our IRT conversions were validated against held-out data of dually administered tests (Fig. 4b) which found that the harmonization pipeline reduced conversion error by 9.5% compared to unadjusted conversions. However, the sample size was small, and we did not have data to independently assess the other two conversions (RAVLT ↔ HVLT and 3. CVLT ↔ HVLT). We attempted to construct a crosswalk for recognition memory trials, but unlike the other items, we could not establish low error IRT results for the recognition item.

Conclusion

Investigators in neuroscience are increasingly turning to Big Data to address replication and reliability issues. However, the aggregation of data from distinct instruments raises new questions about how to integrate data in ways that preserve meaning. This study aggregated data from 53 sites to link scores across common auditory verbal learning tasks (AVLTs). A conversion tool is made freely available online for researchers and clinicians who wish to directly compare memory scores across different instruments. Harmonized AVLT offers opportunities for new, highly powered mega-analytic investigations of verbal learning and memory. This may be

particularly beneficial as a means to functionally characterize the imaging findings from large-scale global open neuroscience initiatives, where interesting imaging features are emerging that are not seen in smaller samples[31].

Data availability

Raw data are available upon reasonable request pending appropriate study approvals and data transfer agreements between participating institutions. Interested researchers should contact Emily Dennis (Emily.dennis@hsc.utah.edu). Code used for analysis and online tool creation are available upon request.

Received: 4 March 2024; Accepted: 12 September 2024
Published online: 16 October 2024

References

1. Thompson, P. M. *et al.* ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry***10**, 100 (2020).
2. Nagaraj, A., Shears, E. & de Vaan, M. Improving data access democratizes and diversifies science. *Proc. Natl. Acad. Sci. U. S. A.***117**, 23490–23498 (2020).
3. Rajtmajer, S. M., Errington, T. M. & Hillary, F. G. How failure to falsify in high-volume science contributes to the replication crisis. *Elife***11**, e78830 (2022).
4. Kennedy, E. *et al.* Harmonizing PTSD severity scales across instruments and sites. *Neuropsychology* Accepted.
5. Pan, F.-F., Huang, L., Chen, K.-L., Zhao, Q.-H. & Guo, Q.-H. A comparative study on the validations of three cognitive screening tests in identifying subtle cognitive decline. *BMC Neurol.***20**, 78 (2020).
6. Roalf, D. R. *et al.* Bridging cognitive screening tests in neurologic disorders: A crosswalk between the short Montreal cognitive assessment and mini-mental state examination. *Alzheimers. Dement.***13**, 947–952 (2017).
7. Woods, S., Delis, D., Scott, J., Kramer, J. & Holdnack, J. The California Verbal Learning Test—second edition: Test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Arch. Clin. Neuropsychol.***21**, 413–420. https://doi.org/10.1016/j.acn.2006.06.002 (2006).
8. Guilmette, T. J. & Rasile, D. Sensitivity, specificity, and diagnostic accuracy of three verbal memory measures in the assessment of mild brain injury. *Neuropsychology***9**, 338–344 (1995).
9. Spreen, O. & Strauss, E. *A Compendium of neuropsychological tests: Administration, norms, and commentary* (Oxford University Press, 1998).
10. Cardenas, S. A., Kassem, L., Brotman, M. A., Leibenluft, E. & McMahon, F. J. Neurocognitive functioning in euthymic patients with bipolar disorder and unaffected relatives: A review of the literature. *Neurosci. Biobehav. Rev.***69**, 193–215 (2016).
11. Jennett, B. & Teasdale, G. Wechsler D. A standardized memory scale for clinical use. *J. Psychol.*
12. Thiruselvam, I. & Hoelzle, J. B. Refined measurement of verbal learning and memory: Application of item response theory to California Verbal Learning Test—Second Edition (CVLT-II) learning trials. *Arch. Clin. Neuropsychol.***35**, 90–104 (2020).
13. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage***208**, 116450 (2020).
14. Radua, J. *et al.* Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. *Neuroimage***218**, 116956 (2020).
15. Wilde, E. A., Dennis, E. L. & Tate, D. F. The ENIGMA brain injury working group: Approach, challenges, and potential benefits. *Brain Imaging Behav.* Under Review, (2019).
16. Coryn, C., Hobson, K. & McCowen, R. Meta-analysis as a method of multi-site evaluation: An example from international development. *Evaluat. J. Austral.***15**, 4–14 (2015).
17. Cifu, D. X. & Dixon, K. J. Chronic effects of neurotrauma consortium. *Brain Inj.***30**, 1397–1398 (2016).
18. Schoenberg, M. R. *et al.* Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Arch. Clin. Neuropsychol.***21**, 693–703 (2006).
19. Benedict, R. H. B., Schretlen, D., Groninger, L. & Brandt, J. Hopkins verbal learning test—revised: Normative data and analysis of inter-form and test-retest reliability. *Clin. Neuropsychol.***12**, 43–55. https://doi.org/10.1076/clin.12.1.43.1726 (1998).
20. Toga, A. W. & Crawford, K. L. The Alzheimer's Disease Neuroimaging Initiative informatics core: A decade in review. *Alzheimers. Dement.***11**, 832–839 (2015).
21. Schmidt, A. *et al.* Structural network disorganization in subjects at clinical high risk for psychosis. *Schizophrenia Bull.*https://doi.org/10.1093/schbul/sbw110 (2016).
22. Lundervold, A. J., Halleland, H. B., Brevik, E. J., Haavik, J. & Sørensen, L. Verbal memory function in intellectually well-functioning adults with ADHD: Relations to working memory and response inhibition. *J. Atten. Disorders***23**, 1188–1198. https://doi.org/10.1177/1087054715580842 (2019).
23. Petrovsky, N. *et al.* Sensorimotor gating is associated with CHRNA3 polymorphisms in schizophrenia and healthy volunteers. *Neuropsychopharmacology***35**, 1429–1439. https://doi.org/10.1038/npp.2010.12 (2010).
24. Broglio, S. P. *et al.* The natural history of sport-related concussion in collegiate athletes: Findings from the NCAA-DoD CARE consortium. *Sports Med.***52**, 403–415 (2022).
25. Logue, M. W. *et al.* Smaller hippocampal volume in posttraumatic stress disorder: A multisite ENIGMA-PGC study: Subcortical volumetry results from posttraumatic stress disorder consortia. *Biol. Psychiatry***83**, 244–253 (2018).
26. Kennedy, E. *et al.* Traumatic brain injury and early onset dementia in post 9–11 veterans. *Brain Inj.***36**, 620–627 (2022).
27. Nasir, N. S. & Hand, V. M. Exploring sociocultural perspectives on race, culture, and learning. *Rev. Educ. Res.***76**, 449–475 (2006).
28. Zopluoglu, C. EstCRM: An R package for Samejima's continuous IRT model. *Appl. Psychol. Meas.***36**, 149 (2012).
29. Shojima, K. A noniterative item parameter solution in each EM cycle of the continuous response model. *Educ. Technol. Res. Dev.***28**, 11–22 (2005).
30. Samejima, F. Homogeneous case of the continuous response model. *Psychometrika***38**, 203–219 (1973).
31. Wang, T. & Zeng, L. Item parameter estimation for a continuous response model using an EM algorithm. *Appl. Psychol. Meas.***22**, 333–344 (1998).
32. Ferrando, P. J. Theoretical and empirical comparisons between two models for continuous item response. *Multivar. Behav. Res.***37**, 521–542 (2002).
33. Priestley, D. R. *et al.* Establishing ground truth in the traumatic brain injury literature: If replication is the answer, then what are the questions?. *Brain Commun.***5**, fcac322 (2023).
34. Naim, M., Katkov, M., Romani, S. & Tsodyks, M. Fundamental law of memory recall. *Phys. Rev. Lett.***124**, 018101 (2020).
35. Barwegen, K., Resch, Z. J., Ovsiew, G. P., Jennette, K. J. & Soble, J. R. A-232 head-to-head comparison of the rey auditory verbal learning test effort score and forced choice embedded performance validity indicators among patients with and without verbal memory impairment. *Arch. Clin. Neuropsychol.***37**, 1388–1388 (2022).
36. Shapiro, A. M., Benedict, R. H., Schretlen, D. & Brandt, J. Construct and concurrent validity of the Hopkins Verbal Learning Test-revised. *Clin. Neuropsychol.***13**, 348–358 (1999).

37. Pugh, M. J. *et al.* Phenotyping the spectrum of traumatic brain injury: A review and pathway to standardization. *J. Neurotrauma***38**, 3222–3234 (2021).

Acknowledgements
Spanish Ministry of Science and Innovation, Instituto de Salud Carlos III: PI15-00852, PI18-00945, JR19-00024, PI17-00481, PI20-00721, Sara Borrell contract (CD19-00149); European Union: NextGenerationEU (PMP21/00051), PI19/01024, Structural Funds, Seventh Framework Program, H2020 Program under the Innovative Medicines Initiative 2 Joint Undertaking: Project PRISM-2 (Grant agreement No.101034377), Project AIMS-2-TRIALS (Grant agreement No 777394), Horizon Europe; National Institutes of Health: U01MH124639, P50MH115846, R01MH113827, R25MH080663, K08MH068540, R01NS100973, R01EB006841, P20GM103472, RO1MH083553, T32MH019535, R01 HD061504, RO1MH083553, R01AG050595, R01AG076838, R01AG060470, R01AG064955, P01AG055367, K23MH095661, R01MH094524, R01MH121246, T32MH019535, R01NS124585, R01NS122827, R61NS120249, R01NS122184, U54EB020403, R01MH116147, R56AG058854, P41EB015922, R01MH111671, P41RR14075, M01RR01066, R01EB006841, R01EB005846, R01 EB000840, RC1MH089257, U24 RR021992, NCRR 5 month-RR001066 (MGH General Clinical Research Center); NSF: 2112455; Madrid Regional Government: B2017/BMD-3740 AGES-CM-2; Dalhousie Medical Research Foundation; Research Nova Scotia, RNS-NHIG-2021-1931; US Department of Defense: Award # AZ150145; US Department of Veterans Affairs: 1I01RX003444; NJ Commission on TBI Research Grants: CBIR11PJT020, CBIR13IRG026: Department of Psychology, University of Oslo; Sunnaas Rehabilitation Hospital: HF F32NS119285; Canadian Institutes of Health Research: Grant 166098; Neurological Foundation of New Zealand; Canterbury Medical Research Foundation, University of Otago. Biogen US, Investigator-initiated grant; Italian Ministry of Health: RF-2019-12370182, Ricerca Corrente RC 23; National Institute on Aging: National Health and Medical Research Council, Investigator Grant APP1176426; PA Health Research: Grant SAP #4100077082 to Dr. Hillary; La Caixa Foundation, ID: 100010434, fellowship code: LCF/BQ/PR22/11920017; Research Council of Norway: 248238; Health Research Council of New Zealand: Sir Charles Hercus Early Career Development (17/039), 14-440; South-Eastern Norway Regional Health Authority, 2018076; Norwegian ExtraFoundation for Health and Rehabilitation: 2015/ FO5146, 2015044; Stiftelsen K.G. Jebsen, SKGJ MED-02; German Research Foundation: DFG grant FOR2107 to Andreas Jansen, JA 1890/7-1, JA 1890/7-2, DFG grant FOR2107 to Igor Nenadić, NE2254/1-2,NE2254/3-1,NE2254/4-1, DFG grant FOR2107, KI588/14-1 and FOR2107, KI588/14-2, DFG, grant FOR2107 DA1151/5-1 and DA1151/5-2, SFB-TRR58, Projects C09 and Z02; Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU); National Health and Medical Research Council: APP1020526; Brain Foundation, Wicking Trust, Collie Trust, Sidney and Fiona Myer Family Foundation. U.S. Army Medical Research and Materiel Command (USAMRMC): Award #13129004; Department of Energy: DE- FG02-99ER62764: Mind Research Network, National Association for Research in Schizophrenia and Affective Disorders: Young Investigator Award, Blowitz Ridgeway and Essel Foundations, NOW ZonMw TOP 91211021, UCLA Easton Clinic for Brain Health, UCLA Brain Injury Research Center, Stan and Patty Silver, Clinical and Translational Research Center: UL1RR033176, UL1TR000124; Mount Sinai Institute for NeuroAIDS Disparities, VA Rehab SPIRE, CDMRP PRAP, VA RR&D IK2RX002922, Veski Fellowship, Femino Foundation grant, Fundación Familia Alonso, Fundación Alicia Koplowitz, CIBERSAM, Madrid Regional Government (B2017/BMD-3740 AGES-CM-2), 2019R1C1C1002457, 21-BR-03-01, 2020M3E5D9079910, 21-BR-03-01, Interdisciplinary Center for Clinical Research (IZKF) of the medical faculty of Münster.

Author contributions
Original data were collected by the members of the ENIGMA Clinical Endpoints Working Group. E.K., H.M.L., P.W.L., M.J.P., P.M.T., F.G.H., D.F.T., E.A.W., and E.L.D. conceived of the project. E.L.D. and H.M.L. compiled and cleaned the data. E.K. and S.V. completed the processing and analysis. E.K., H.M.L., P.W.L., and E.L.D. wrote the manuscript and all authors reviewed, edited, and approved the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Ethical standards
The overall study was approved by the institutional review board of the University of Utah. This study was limited to secondary deidentified data analysis, and was designated as non-human subjects research.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-72968-x.

Correspondence and requests for materials should be addressed to E.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024

for the ENIGMA Clinical Endpoints Working Group

Eamonn Kennedy[1,2,3], Shashank Vadlamani[1], Hannah M. Lindsey[1,3], Pui-Wa Lei[4], Mary Jo-Pugh[1,2], Maheen Adamson[10,11], Martin Alda[12], Silvia Alonso-Lana[13,14,15], Sonia Ambrogi[16], Tim J. Anderson[17,18,19], Celso Arango[20], Robert F. Asarnow[21,22,23], Mihai Avram[24], Rosa Ayesa-Arriola[14,25], Talin Babikian[21,26], Nerisa Banaj[16], Laura J. Bird[27], Stefan Borgwardt[24,28], Amy Brodtmann[29,30], Katharina Brosch[31], Karen Caeyenberghs[32], Vince D. Calhoun[33], Nancy D. Chiaravalloti[34,35], David X. Cifu[36], Benedicto Crespo-Facorro[14,37], John C. Dalrymple-Alford[17,18,38], Kristen Dams-O'Connor[39,40], Udo Dannlowski[41], David Darby[42,43,44], Nicholas Davenport[45,46], John DeLuca[35,47], Covadonga M. Diaz-Caneja[20], Seth G. Disner[45,46], Ekaterina Dobryakova[35,48], Stefan Ehrlich[49,50], Carrie Esopenko[39], Fabio Ferrarelli[51], Lea E. Frank[52], Carol Franz[53,54], Paola Fuentes-Claramonte[13,14], Helen Genova[35,55], Christopher C. Giza[26,56,57], Janik Goltermann[41], Dominik Grotegerd[41], Marius Gruber[41,58], Alfonso Gutierrez-Zotes[14,59,60], Minji Ha[61], Jan Haavik[62,63], Charles Hinkin[21], Kristen R. Hoskinson[64,65], Daniela Hubl[66], Andrei Irimia[67,68,69], Andreas Jansen[31], Michael Kaess[70,71], Xiaojian Kang[10], Kimbra Kenney[72,73], Barbora Keřková, Mohamed Salah Khlif[29], Minah Kim[74,75], Jochen Kindler[70], Tilo Kircher[31], Karolina Knížková[76,77], Knut K. Kolskår[78,79,80], Denise Krch[35,48], William S. Kremen[53,54], Taylor Kuhn[21], Veena Kumari[81], Jun Soo Kwon[61,74,75], Roberto Langella[16], Sarah Laskowitz[82], Jungha Lee[61], Jean Lengenfelder[35,48], Spencer W. Liebel[1,3], Victoria Liou-Johnson[10], Sara M. Lippa[73,83], Marianne Løvstad[79,80], Astri Lundervold[84], Cassandra Marotta[42,43], Craig A. Marquardt[45,46], Paulo Mattos[85], Ahmad Mayeli[51], Carrie R. McDonald[86,87], Susanne Meinert[41,88], Tracy R. Melzer[17,18,38], Jessica Merchán-Naranjo[20], Chantal Michel[70], Rajendra A. Morey[82,89], Benson Mwangi[90], Daniel J. Myall[18], Igor Nenadić[31], Mary R. Newsome[91,92], Abraham Nunes[12,93], Terence O'Brien[94,95], Viola Oertel[96], John Ollinger[73], Alexander Olsen[97,98,99], Victor Ortiz García Foz[25], Mustafa Ozmen[100], Heath Pardoe[44], Marise Parent[101], Fabrizio Piras[16], Federica Piras[16], Edith Pomarol-Clotet[14], Jonathan Repple[41,58], Geneviève Richard[78], Jonathan Rodriguez[53], Mabel Rodriguez[76], Kelly Rootes-Murdy[33], Jared Rowland[102,103,104], Nicholas P. Ryan[32,105], Raymond Salvador[14], Anne-Marthe Sanders[78,79,80], Andre Schmidt[106], Jair C. Soares[90], Gianfranco Spalletta[16], Filip Španiel[76,107], Alena Stasenko[53,87], Frederike Stein[31], Benjamin Straube[31], April Thames[21], Florian Thomas-Odenthal[31], Sophia I. Thomopoulos[5], Erin Tone[108], Ivan Torres[109,110], Maya Troyanskaya[91,92], Jessica A. Turner[111], Kristine M. Ulrichsen[78,79,80], Guillermo Umpierrez[112], Elisabet Vilella[14,59,60], Lucy Vivash[42,43], William C. Walker[113,114], Emilio Werden[44], Lars T. Westlye[78,79,115], Krista Wild[116], Adrian Wroblewski[31], Mon-Ju Wu[90], Glenn R. Wylie[35,117], Lakshmi N. Yatham[109], Giovana B. Zunta-Soares[90], Paul M. Thompson[5,6], David F. Tate[1,3], Frank G. Hillary[7,8,9], Emily L. Dennis[1,3] & Elisabeth A. Wilde[1,3]

[10]WRIISC-WOMEN & Rehabilitation Department, VA Palo Alto, Palo Alto, CA, USA. [11]Neurosurgery, Stanford School of Medicine, Stanford, CA, USA. [12]Department of Psychiatry, Dalhousie University, Halifax, Canada. [13]FIDMAG Research Foundation, Barcelona, Spain. [14]Centro Investigación Biomédica en Red Salud Mental (CIBERSAM), Madrid, Spain. [15]Ace Alzheimer Center Barcelona - Universitat Internacional de Catalunya, Barcelona, Spain. [16]Laboratory of Neuropsychiatry, Santa Lucia Foundation IRCCS, Rome, Italy. [17]Department of Medicine, University of Otago, Christchurch, New Zealand. [18]New Zealand Brain Research Institute, Christchurch, New Zealand. [19]Department of Neurology, Te Whatu Ora – Health New Zealand Waitaha Canterbury, Christchurch, New Zealand. [20]Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, School of Medicine, Universidad Complutense, Madrid, Spain. [21]Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and

Human Behavior, UCLA, Los Angeles, CA, USA. [22]Brain Research Institute, UCLA, Los Angeles, CA, USA. [23]Department of Psychology, UCLA, Los Angeles, CA, USA. [24]Translational Psychiatry, Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany. [25]Department of Psychiatry, Marqués de Valdecilla University Hospital, IDIVAL, School of Medicine, University of Cantabria, Santander, Spain. [26]UCLA Steve Tisch BrainSPORT Program, Los Angeles, CA, USA. [27]Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, Melbourne, VIC, Australia. [28]Center of Brain, Behaviour and Metabolism (CBBM), University of Lübeck, Lübeck, Germany. [29]Cognitive Health Initiative, Central Clinical School, Monash University, Melbourne, VIC, Australia. [30]Department of Medicine, Royal Melbourne Hospital, Melbourne, VIC, Australia. [31]Department of Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany. [32]Cognitive Neuroscience Unit, School of Psychology, Deakin University, Geelong, Australia. [33]Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, and Emory University University, Atlanta, GA, USA. [34]Centers for Neuropsychology, Neuroscience & Traumatic Brain Injury Research, Kessler Foundation, East Hanover, NJ, USA. [35]Department of Physical Medicine & Rehabilitation, New Jersey Medical School, RutgersNewark, NJ, USA. [36]Rehabilitation Medicine Department, National Institutes of Health Clinical Center, Bethesda, MD, USA. [37]Department of Psychiatry, Virgen del Rocio University Hospital, School of Medicine, University of Seville, IBIS, Seville, Spain. [38]School of Psychology, Speech and Hearing, University of Canterbury, Christchurch, New Zealand. [39]Department of Rehabilitation and Human Performance, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [40]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [41]Institute for Translational Psychiatry, University of Münster, Münster, Germany. [42]Department of Neuroscience, Monash University, Melbourne, Australia. [43]Department of Neurology, Alfred Health, Melbourne, Australia. [44]The Florey Institute of Neuroscience and Mental Health, Melbourne, Australia. [45]Department of Psychiatry and Behavioral Sciences, University of Minnesota Medical School, Minneapolis, MN, USA. [46]Minneapolis VA Health Care System, Minneapolis, MN, USA. [47]Kessler Foundation, East Hanover, NJ, USA. [48]Center for Traumatic Brain Injury, Kessler Foundation, East Hanover, NJ, USA. [49]Translational Developmental Neuroscience Section, Division of Psychological and Social Medicine and Developmental Neurosciences, Faculty of Medicine, TU Dresden, Dresden, Germany. [50]Eating Disorders Research and Treatment Center, Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Dresden, Germany. [51]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA. [52]Department of Psychology, University of Oregon, Eugene, OR, USA. [53]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. [54]Center for Behavior Genetics of Aging, University of California, San Diego, La Jolla, CA, USA. [55]Center for Autism Research, Kessler Foundation, East Hanover, NJ, USA. [56]Department of Pediatrics, Division of Neurology, UCLA Mattel Children's Hospital, Los Angeles, CA, USA. [57]Department of Neurosurgery, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. [58]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Goethe University, Frankfurt, Germany. [59]Hospital Universitari Institut Pere Mata, Tarragona, Spain. [60]Institut d'Investiació Sanitària Pere Virgili-CERCA, Universitat Rovira i Virgili, Tarragona, Spain. [61]Department of Brain and Cognitive Sciences, Seoul National University College of Natural Sciences, Seoul, South Korea. [62]Department of Biomedicine, University of Bergen, Bergen, Norway. [63]Division of Psychiatry, Haukeland University Hospital, Bergen, Norway. [64]Center for Biobehavioral Health, The Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, USA. [65]Section of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA. [66]Translational Research Centre, University Hospital of Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland. [67]Ethel Percy Andrus Gerontology Center, Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, USA. [68]Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA. [69]Department of Quantitative & Computational Biology, Dornsife College of Arts & Sciences, University of Southern California, Los Angeles, CA, USA. [70]University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland. [71]Clinic of Child and Adolescent Psychiatry, Centre of Psychosocial Medicine, University of Heidelberg, Heidelberg, Germany. [72]Department of Neurology, Uniformed Services University, Bethesda, MD, USA. [73]National Intrepid Center of Excellence, Walter Reed National Military Medical Center, Bethesda, MD, USA. [74]Department of Neuropsychiatry, Seoul National University Hospital, Seoul, South Korea. [75]Department of Psychiatry, Seoul National University College of Medicine, Seoul, South Korea. [76]National Institute of Mental Health, Klecany, Czech Republic. [77]Department of Psychiatry, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic. [78]NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [79]Department of Psychology, University of Oslo, Oslo, Norway. [80]Sunnaas Rehabilitation Hospital, Nesodden, Norway. [81]Department of Life Sciences, College of Health, Medicine and Life Sciences, Brunel University London, Uxbridge, UK. [82]Brain Imaging and Analysis Center, Duke University, Durham, NC, USA. [83]Department of Neuroscience, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. [84]Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway. [85]Institute D'Or for Research and Education (IDOR), São Paulo, Brazil. [86]Department of Radiation Medicine and Applied Sciences and Psychiatry, UC San Diego, La Jolla, CA, USA. [87]Center for Multimodal Imaging and Genetics, UC San Diego, San Diego, CA, USA. [88]Institute for Translational Neuroscience, University of Münster, Münster, Germany. [89]VISN 6 MIRECC, Durham VA, Durham, NC, USA. [90]Center of Excellence On Mood Disorders, Louis A. Faillace, MD, Department of Psychiatry and Behavioral Sciences, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA. [91]Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA. [92]H. Ben Taub Department of Physical Medicine and Rehabilitation, Baylor College of Medicine, Houston, TX, USA. [93]Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada. [94]Department of Medicine, The Royal Melbourne Hospital, The University of Melbourne, Parkville, VIC, Australia. [95]Department of Neuroscience, The Central Clinical School, Alfred Health, Monash University, Melbourne, VIC, Australia. [96]Department of Psychiatry, Psychosomatic Medicine and Psychotherapiey, Frankfurt University, Frankfurt, Germany. [97]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway. [98]Department of Physical

Medicine and Rehabilitation, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [99]NorHEAD - Norwegian Centre for Headache Research, Trondheim, Norway. [100]Department of Electrical and Electronics Engineering, Antalya Bilim University, Antalya, Turkey. [101]Neuroscience Institute & Department of Psychology, Georgia State University, Atlanta, GA, USA. [102]W.G. (Bill) Hefner VA Medical Center, Salisbury, NC, USA. [103]Department of Neurobiology & Anatomy, Wake Forest School of Medicine, Winston-Salem, NC, USA. [104]VA Mid-Atlantic Mental Illness Research Education and Clinical Center (MA-MIRECC), Durham, NC, USA. [105]Department of Paediatrics, The University of Melbourne, Parkville, VIC, Australia. [106]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland. [107]3Rd Faculty of Medicine, Charles University, Prague, Czech Republic. [108]Department of Psychology, Georgia State University, Atlanta, GA, USA. [109]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada. [110]British Columbia Mental Health and Substance Use Services Research Institute, Vancouver, BC, Canada. [111]Psychiatry and Behavioral Health, Ohio State Wexner Medical Center, Columbus, OH, USA. [112]Division of Endocrinology, Emory University School of Medicine, Atlanta, GA, USA. [113]Department of Physical Medicine and Rehabilitation, Virginia Commonwealth University, Richmond, VA, USA. [114]Richmond Veterans Affairs (VA) Medical Center, Central Virginia VA Health Care System, Richmond, VA, USA. [115]KG Jebsen Center for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway. [116]Department of Psychology, Phoenix VA Health Care System, Phoenix, AZ, USA. [117]Rocco Ortenzio Neuroimaging Center, Kessler Foundation, East Hanover, NJ, USA.