

Harmonizing PTSD Severity Scales Across Instruments and Sites

Eamonn Kennedy^{1, 2, 3}, Emily L. Dennis^{1, 2}, Hannah M. Lindsey^{1, 2}, Terri deRoos-Cassini⁴, Stefan Du Plessis^{5, 6}, Negar Fani⁷, Milissa L. Kaufman^{8, 9}, Nastassja Koen^{5, 6}, Christine L. Larson¹⁰, Sarah Laskowitz¹¹, Lauren A. M. Lebois^{8, 12}, Rajendra A. Morey^{11, 13}, Mary R. Newsome^{14, 15}, Cori Palermo^{8, 12}, Nicholas J. Pastorek^{14, 15}, Abigail Powers⁷, Randall Scheibel^{14, 15}, Soraya Seedat¹⁶, Antonia Seligowski^{8, 12}, Dan J. Stein^{5, 6}, Jennifer Stevens⁷, Delin Sun^{11, 13}, Paul Thompson¹⁷, Maya Troyanskaya^{14, 15}, Sanne J. H. van Rooij⁷, Amanda A. Watts¹¹, Carissa W. Tomas¹⁸, Wright Williams¹⁵, Frank G. Hillary^{19, 20}, Mary Jo Pugh^{1, 2, 3}, Elisabeth A. Wilde^{1, 2, 14}, and David F. Tate^{1, 2}

¹ Department of Neurology, University of Utah School of Medicine

² George E. Wahlen Veterans Affairs Medical Center, Salt Lake City, Utah, United States

³ Division of Epidemiology, University of Utah

⁴ Department of Surgery, Division of Trauma and Acute Care Surgery and Comprehensive Injury Center, Medical College of Wisconsin

⁵ Department of Psychiatry and Mental Health, University of Cape Town

⁶ SAMRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry and Mental Health and Neuroscience Institute, University of Cape Town

Editor's Note. Keith Owen Yeates served as the action editor for this article.—KOY

This article was published Online First July 7, 2022.

Eamonn Kennedy  <https://orcid.org/0000-0002-9211-1271>

This work was supported by the National Institutes of Health (NIH) National Institute of Neurological Disorders and Stroke, award #R61NS120249, and the Long-Term Impact of Military Relevant Brain Injury Consortium (LIMBIC) Award W81XWH18PH/TBIRPLIMBIC under Awards No. W81XWH1920067 and W81XWH1320095. Secondary data provided for the mega analysis were supported by: Award number R21MH112956, the Anonymous Women's Health Fund, Kasparian Fund, Trauma Scholars Fund; The South African Medical Research Council for the "Shared Roots" Flagship Project, Grant MRC-RFA-IFSP-01-2013/SHARED ROOTS" through funding received from the South African National Treasury under its Economic Competitiveness and Support Package; Funding from the SAMRC Unit on Risk & Resilience in Mental Disorders; The Narsad Young Investigators award; The PA Health Research Grant SAP #4100077082 to Frank G. Hillary; Award numbers: (NIH) K23MH125920; AHA 20CDA35310031; MH098212; MH071537; M01RR00039; UL1TR000454; HD071982; HD085850; U54 EB020403; R01 MH116147; R56 AG058854; P41 EB015922; R01 MH111671; F32MH109274; and K23AT009713. All contents and opinions expressed are sole responsibility of the authors and do not necessarily represent the official views of any funding sources.

Eamonn Kennedy played lead role in formal analysis, investigation, methodology, software, validation, visualization, writing of original draft, and writing of review and editing and equal role in conceptualization and data curation. Emily L. Dennis played equal role in data curation, project administration, supervision, and writing of review and editing. Hannah M. Lindsey played equal role in data curation and writing of review and editing. Terri deRoos-Cassini played equal role in data curation, investigation, and writing of review and editing. Stefan Du Plessis played equal role in data curation, investigation, and writing of review and editing. Negar Fani played equal role in data curation, investigation, and writing of review and editing. Milissa L. Kaufman played equal role in data curation, investigation, and writing of review and editing. Nastassja Koen played equal role in data curation, investigation, and writing of review and editing. Christine L. Larson played equal role in data curation, investigation, and writing of

review and editing. Sarah Laskowitz played equal role in data curation, investigation, and writing of review and editing. Lauren A. M. Lebois played equal role in data curation, investigation, and writing of review and editing. Rajendra A. Morey played equal role in data curation and investigation. Mary R. Newsome played equal role in data curation, investigation, and writing of review and editing. Cori Palermo played equal role in data curation, investigation, and writing of review and editing. Nicholas J. Pastorek played equal role in data curation, investigation, and writing of review and editing. Abigail Powers played equal role in data curation, investigation, and writing of review and editing. Randall Scheibel played equal role in data curation, investigation, methodology, validation, and writing of review and editing. Soraya Seedat played equal role in data curation, investigation, and writing of review and editing. Antonia Seligowski played equal role in data curation, investigation, and writing of review and editing. Dan J. Stein played equal role in data curation, investigation, and writing of review and editing. Jennifer Stevens played equal role in data curation, investigation, and writing of review and editing. Delin Sun played equal role in data curation, investigation, and writing of review and editing. Paul Thompson played equal role in data curation, project administration, and writing of review and editing. Maya Troyanskaya played equal role in data curation, formal analysis, investigation, methodology, validation, and writing of review and editing. Sanne J. H. van Rooij played equal role in data curation, investigation, and writing of review and editing. Amanda A. Watts played equal role in data curation, project administration, and writing of review and editing. Carissa W. Tomas played equal role in data curation, investigation, and writing of review and editing. Wright Williams played equal role in data curation, investigation, and writing of review and editing. Frank G. Hillary played equal role in conceptualization, data curation, investigation, supervision, and writing of review and editing. Mary Jo Pugh played equal role in investigation, resources, supervision, and writing of review and editing. Elisabeth A. Wilde played lead role in supervision and equal role in conceptualization, funding acquisition, investigation, project administration, and writing of review and editing. David F. Tate played lead role in supervision and equal role in conceptualization, funding acquisition, project administration, and writing of review and editing.

Correspondence concerning this article should be addressed to Eamonn Kennedy, Department of Neurology, University of Utah School of Medicine, Salt Lake City, UT 84108, United States. Email: eamonn.kennedy@utah.edu

⁷ Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine

⁸ Department of Psychiatry, Harvard Medical School

⁹ Division of Women's Mental Health, McLean Hospital, Belmont, Massachusetts, United States

¹⁰ Department of Psychology, University of Wisconsin-Milwaukee

¹¹ Brain Imaging and Analysis Center, Duke University

¹² Division of Depression and Anxiety Disorders, McLean Hospital, Belmont, Massachusetts, United States

¹³ VISN 6 MIRECC, Durham VA, Durham, North Carolina, United States

¹⁴ H. Ben Taub Department of Physical Medicine and Rehabilitation, Baylor College of Medicine

¹⁵ Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas, United States

¹⁶ SU/UCT MRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry, Stellenbosch University

¹⁷ Imaging Genetics Center, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, Los Angeles, California, United States

¹⁸ Department of Epidemiology and Comprehensive Injury Center, Medical College of Wisconsin

¹⁹ Department of Neurology, Hershey Medical Center, State College

²⁰ Department of Psychology, Penn State University

Objective: The variety of instruments used to assess posttraumatic stress disorder (PTSD) allows for flexibility, but also creates challenges for data synthesis. The objective of this work was to use a multisite mega analysis to derive quantitative recommendations for equating scores across measures of PTSD severity. **Method:** Empirical Bayes harmonization and linear models were used to describe and mitigate site and covariate effects. Quadratic models for converting scores across PTSD assessments were constructed using bootstrapping and tested on hold out data. **Results:** We aggregated 17 data sources and compiled an $n = 5,634$ sample of individuals who were assessed for PTSD symptoms. We confirmed our hypothesis that harmonization and covariate adjustments would significantly improve inference of scores across instruments. Harmonization significantly reduced cross-dataset variance (28%, $p < .001$), and models for converting scores across instruments were well fit (median $R^2 = 0.985$) with an average root mean squared error of 1.46 on sum scores. **Conclusions:** These methods allow PTSD symptom severity to be placed on multiple scales and offers interesting empirical perspectives on the role of harmonization in the behavioral sciences.

Key Points

Question: The precise relationship between scores on different PTSD assessments remains unclear because it is hard to isolate the effects of instrumentation in practice. **Findings:** We found that individual data sources each come with distinct characteristics and biases that must be addressed before the relationship between different PTSD assessments can be observed and modeled independently. **Importance:** We propose methods that allow scores on different measures of PTSD symptom severity to be compared, which may reduce burden on patients, participants, and researchers. **Next Steps:** Future work could use these ideas to situate the results of new studies within the larger body of historical literature.

Keywords: harmonization, PTSD, screening instruments, mega analysis

Large-scale data sharing initiatives offer opportunities to improve robustness by synthesizing multiple data sources (Thompson et al., 2020). However, in the behavioral sciences, differences in psychometric evaluation can confound the aggregation of data (Houtkoop et al., 2018; Towse et al., 2021). For example, researchers and clinicians can select from a variety of instruments for measuring posttraumatic stress disorder (PTSD), which is a prevalent and burdensome mental health condition (Kessler et al., 2005; Norris & Hamblen, 2004). Instruments that assess PTSD symptom severity broadly classify into three groups: (a) clinical interviews such as the Clinician-Administered PTSD Scale for *Diagnostic and Statistical Manual of Mental Disorders-5* (DSM-5; American Psychiatric Association, 2013), abbreviated CAPS-5 (Weathers et al., 2013, 2018); (b) brief self-assessments such as the Davidson Trauma Scale (DTS) which can briefly screen for provisional diagnosis (Davidson et al., 1997), and (c) thematically specific severity scales which are designed to assess a particular group (Wilkins et al., 2011).

This variety of assessments affords flexibility, but also creates challenges for data synthesis. Even within one assessment, test

instructions and items are continually modified to match advancing diagnostic criteria. For example, the DSM-5 criteria for PTSD (American Psychiatric Association, 2013) are distinct from previous iterations (Brett et al., 1988), and different factor solutions have been proposed (Shelby et al., 2005).

Establishing standards for converting scores across PTSD symptom inventories could improve clinical and research consistency. However, it is challenging to isolate instrumental effects because severity scores depend on at least five factors: (a) clinical features and presentation; (b) intrinsic biological variables such as age; (c) distinct procedures across studies and sites; (d) instrumental variations, such as distinct question phrasings; and (e) statistical error and randomness. To accurately convert scores, the instrumental component must be isolated from other sources of variation, but data from single sources is typically subject to specific biases since most studies recruit and sample for specific conditions or traits (Radua et al., 2020; Pugh et al., 2021). Therefore, a secondary mega analysis is a good solution for identifying and removing unwanted effects (Boedhoe et al., 2019).

We report a multisite ($n = 17$ datasets) mega study analysis of five common instruments used for PTSD assessment. We leveraged recent data harmonization algorithms (Pomponio et al., 2020) to remove site effects. The component of severity scores associated with instrumentation was isolated from covariate effects, and calculations for converting scores across measures were tested on hold out data (data not used during model construction). Our main hypothesis was that without corrections, percentage and percentile models (Kolen & Brennan, 2004) would be confounded, while performance would be significantly improved by the harmonization of data sources, covariate adjustments, and by models tolerant of some nonlinearity across instruments.

Method

Data Sources

This secondary mega analysis draws from a range of military and civilian studies. We petitioned collaborators for item level data, drawing from the psychiatric genomics consortium and the enhancing neuroimaging genetics through meta-analysis consortium (PGC-ENIGMA) PTSD working group (Logue et al., 2018), the ENIGMA brain injury working group (Wilde et al., 2021), and the long-term impact of military-relevant brain injury consortium—chronic effects of neurotrauma consortium (LIMBIC-CENC; Cifu & Dixon, 2016). We obtained 17 datasets that performed different combinations of PTSD assessments. Data quality and consistency were confirmed during discussions among authors who performed the primary data collection. All assessments were conducted in English. The University of Utah provided overall institutional review board (IRB) study approval and data use agreements for the following sources:

1. DOD-ADNI: a study of brain aging in Vietnam war veterans (Weiner et al., 2017).
2. iSCORE: the imaging support for the study of cognitive rehabilitation (Tate et al., 2019).
3. PT: the personality traits and brain matter aberrations as potential markers of mTBI and PTSD study.
4. CE: the longitudinal study of chronic effects of traumatic brain injury (TBI) in Veterans and service members.
5. Blast I: Blast I is an fMRI study of TBI associated with blast injury.
6. Blast II: Blast II is a renewal of the initial study capturing similar data.
7. NBS-DoD: neural and behavioral sequelae of blast-related traumatic brain injury.
8. SARChI: Stellenbosch's South African research chairs initiative (Suliman et al., 2014).
9. GTP: the Grady trauma project (Gillespie et al., 2009).
10. iSTAR: imaging study of trauma and resilience (Weis et al., 2022).
11. TSS: the McLean trauma spectrum study (Lebois et al., 2021).

12. NEST: the McLean neurocardiac effects of stress and trauma study.
13. DCHS: the Drakenstein child health study of prenatal mothers (Donald et al., 2018).
14. VCTP: neuroimaging meditation therapy in Veterans with comorbid mild TBI and PTSD.
15. MC: a military mission connect study.
16. MIRECC-DU: mental illness research education clinical, centers of excellence, Duke University.
17. MIRECC-D: mental illness research education clinical, centers of excellence, Durham Veterans Affairs (Sun et al., 2020).

Additionally, data from LIMBIC-CENC (Cifu & Dixon, 2016) were held out during model construction to enable demonstration and independent testing. CE, Blast I/II, and some of the NBS-DoD data were acquired using very similar inclusion/exclusion criteria, but collection was stratified over time.

Inclusion Criteria

Adults aged over 18 years who completed at least one assessment to a level of <20% missingness were included. PTSD severity screeners were obtained at first entry to care facilities or at the initiation of research studies. Any repeated measurements per person were excluded, alongside measurements after interventions. Since the number of nonsymptom reporting cases is largely dependent on the inclusion criteria of studies, individuals who reported the lowest possible sum score (e.g., a total severity score of zero on PTSD Checklist for DSM-5 [PCL-5]) were excluded (see Limitations section).

Measures and Characteristics

PTSD inventories are typically designed to elicit one item level response per diagnostic criteria. This means 20 unique items were recorded for DSM-5 and 17 unique items were recorded for Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) assessments. All instruments included overlapping items that facilitated harmonization.

The PCL-5 is a 20-item self-report measure that assesses the 20 DSM-5 symptoms of PTSD using a 5-point Likert scale ranging from 0 to 4 (*not at all* to *extremely*). The checklist asks the participant to consider the level of symptom severity over the last month. The PCL-5 severity score is the sum of all items scores and ranges from 0 to 80 (Weathers et al., 1993, Wortmann et al., 2016).

The PTSD Checklist for DSM-IV (PCL-C) is a 17-item self-report measure that assesses the severity of the 17 DSM-IV symptoms of PTSD using a 5-point Likert scale ranging from 1 to 5 (*not at all* to *extremely*) as experienced over the last month. The PCL-C severity score is the sum of all items, which ranges from 17 (no symptoms) to 85 (Weathers et al., 1993).

The PTSD Checklist for DSM-IV—Military Version (PCL-M). Like the PCL-C, the PCL-M is a 17-item self-report that measures the severity of the 17 DSM-IV PTSD symptoms over the last month on a 5-point Likert scale ranging from 1 to 5 (*not at all* to *extremely*). The items of the PCL-M are the same as the PCL-C except that the

PCL-M wording addresses a stressful military experience. A prior synthesis of the PCL-C and PCL-M demonstrated strong consistency, reliability, and convergence (Wilkins et al., 2011), and in this work we denote both DSM-IV PCL screeners as “PCL”. Item level military/civilian differences are described in detail (see Results section).

The *Davidson Trauma Scale (DTS)* is a 17-item self-report measure that assesses the frequency and severity of the 17 DSM-IV symptoms of PTSD. Each item of the DTS severity scale ranges from (0 = *not at all distressing* to 4 = *extremely distressing*) and the responder is asked to consider symptoms within the last week. The total sum score ranges from 0 to 136. We considered only the severity scale of DTS which ranges from 0 to 68 (Davidson et al., 1997).

The *modified PTSD Symptom Scale (mPSS)* is a 17-item self-report measure that asks about how upsetting the 17 DSM-IV symptoms of PTSD severity have been within the last 2 weeks, with items rated on 4-point Likert scale consisting of 0 (*not at all*), 1 (*once per week or less/a little bit/once in a while*), 2 (*two to four times per week/somewhat/half the time*) to 3 (*five or more times per week/very much/almost*). Sum severity scores range from 0 to 51 (Falsetti et al., 1993).

Covariates of age, sex/gender, site/study, and military/civilian status were included. PTSD-associated conditions such as substance use disorder, depression, and early onset cognitive impairment (Kennedy et al., 2022) were not considered in the primary analysis as they were not recorded consistently across studies. Race/ethnicity characteristics were also not recorded consistently across studies. Military status was broadly defined, and included U.S. Veterans of the Vietnam war as well as Veterans of operation enduring freedom/operation Iraqi freedom/operation new dawn (OEF/OIF/OND). While some studies recorded gender, others recorded biological sex, so these characteristics were aggregated into a single sex/gender variable.

Statistical Analysis

Analysis was performed in Python 3. Kruskal–Wallis H tests (omnibus) were used to test for significance differences across groups. Welch’s t tests were used for post hoc pairwise comparisons. Where multiple tests were performed, q statistics were calculated at a threshold of 0.05 to reduce false discovery rates. Absolute severity scores were converted to fraction/percentage severity scores by subtracting the minimum assessment score from each observed score and dividing by the full range of the instrument.

The aim of harmonization was to remove unwanted site effects while preserving instrumental effects for further analysis. If absolute scores were harmonized, this would remove both site effects and also the absolute differences between instruments (e.g., baseline offsets between scales). Therefore, we used the ComBat-GAM algorithm (Pomponio et al., 2020) to harmonize percentage scores over all datasets. This method explicitly protected covariates and instrumental effects. After harmonization, percentage severity scores were returned to absolute scores on their respective instruments with site effects removed. Harmonization efficacy was measured as the reduction in cross-dataset variance when comparing raw scores to their postharmonized equivalents. Coefficients of

determination (R^2) were used to calculate the deviation of data from models. An R^2 of one means a model perfectly fits the data.

Partitioning, Training, and Model Description

Ordinary least squares (OLS) linear models were used to adjust for covariates and to convert scores across instruments. After removing covariate effects, the square of sum scores was used as an additional term during instrument conversion to capture potential nonlinearities across assessments. Comparing random subsamples of the data reduced bias associated with variations in clinical presentation, and allowed for a measurement of confidence on the inferred scores (Choi et al., 2014). The quantity of interest is a predicted line of model fit, so we measured confidence using root mean squared error (RMSE). Since the number of observations differed for each assessment, we elected to bootstrap using $N = \text{argmin}([n_A, n_B])/2$ samples in each model fit iteration, where n_A and n_B and the total observations of assessment A and B , respectively. This means the training fraction had an upper bound of 50%. Summary results were determined by averaging over the coefficients recovered from all model iterations.

Transparency and Openness

Raw data is available upon reasonable request pending study approval and data transfer agreements between all participating institutions. Codes used for analysis can be provided by the authors upon reasonable request.

Results

Data Summary

The summary statistics of the 17 datasets (see section Method) are shown in Table 1. The total sample was $n = 6,771$ but this reduced to $n = 5,634$ after exclusions were applied. The median age was 36 years old with an interquartile range of 28–46 years, and 36.4% of all study participants were female. The data comprised seven civilian studies, nine military studies, and one mixed population study. The total counts per assessment were (a) PCL-5: $n = 1,325$, (b) PCL-C/M: $n = 786$, (c) DTS: $n = 3,196$, and (d) mPSS: $n = 327$.

Figure 1 provides an overview of symptom reporting across all datasets. Figure 1a plots the histogram of all percentage severity scores. Figure 1b shows percentage sum scores broken out as boxplots by military status and sex/gender. An omnibus test confirmed significant differences between the groups in Figure 1b. Female civilians reported the most severe symptoms overall ($p < .001$). Figure 1c shows the mean severity for each item across all datasets set to the same scale (range: [0:4] as per PCL-5). Sleep disturbance, hypervigilance, and negative feelings were the most intensely reported symptoms. Military status was associated with elevated risk taking (Figure 1c—right). Civilian status was associated with increased report of upsetting reminders, and increased avoidance of upsetting reminders and feelings (Figure 1c—right).

Unadjusted Conversion

We first consider the conversion from PCL-5 to PCL severity scores. PCL-5 sum scores range from (0, 80), while PCL sum scores range from (17, 85). Therefore, the line where the percentage

Table 1
Summary Statistics for the 17 Datasets

Dataset	Population	<i>N</i>	Female	Male	Median age	Measure	Mean score
VA-MIRECC-D	Military	2,988	0.22	0.78	36	DTS	25.4
VA-MIRECC-DU	Military	208	0.19	0.81	39	DTS	20.8
DCHS	Civilian	327	1	0	26	MPSS	23.1
SARChI	Civilian	591	0.71	0.29	44	PCL-5	33
GTP	Civilian	335	0.86	0.14	44	PCL-5	32.4
iSTAR	Civilian	180	0.58	0.42	33	PCL-5	25.1
TSS	Civilian	135	0.83	0.17	35	PCL-5	49.1
NEST	Civilian	15	0.8	0.2	33	PCL-5	37.4
PT	Military	60	0.1	0.9	38	PCL-5	43.8
VCTP	Military	9	0.33	0.67	38	PCL-5	47.6
CE	Military	75	0.08	0.92	33	PCL-C	44.4
Blast II	Military	50	0.02	0.98	30	PCL-C	51.5
Blast I	Military	47	0.11	0.89	31	PCL-C	39.6
NBS-DoD	Mixed	31	0.29	0.71	29	PCL-C	36.1
DoD-ADNI	Military	242	0	1	69	PCL-M	10.8*
iSCORE	Military	195	0.12	0.88	34	PCL-M	45.7
MC	Civilian	146	0.28	0.72	27	PCL-C	27.3

Note. VA-MIRECC-D = Veterans Affairs - Mental Illness Research Education Clinical, Centers of Excellence Durham; VA-MIRECC-DU = Veterans Affairs - Mental Illness Research Education Clinical, Centers of Excellence, Duke University; DCHS = Drakenstein child health study; MPSS = Modified PTSD Symptom Scale; SARChI = Stellenbosch's South African research chairs initiative; GTP = Grady trauma project; iSTAR = imaging study of trauma and resilience; TSS = trauma spectrum study; NEST = neurocardiac effects of stress and trauma; PT = personality traits; VCTP = neuroimaging meditation therapy in Veterans with comorbid mild TBI and PTSD; MC = military mission connect study; DTS = Davidson Trauma Scale; CE = chronic effects; NBS-DoD = neural and behavioral sequelae of blast-related traumatic brain injury; DOD-ADNI = study of brain aging in Vietnam war veterans; Blast I = fMRI study of TBI associated with blast injury; Blast II = renewal of the initial study capturing similar data; iSCORE = imaging support for the study of cognitive rehabilitation; PCL-5 = PTSD checklist for DSM-5; PCL-C = PTSD checklist for DSM-IV; PCL-M = PTSD Checklist for DSM-IV—Military Version. Statistics for each of the datasets are shown after exclusion criteria were applied (see Method section).

* indicates each item was recorded on a binary scale instead of a Likert scale, which results in an apparently low average sum score.

severity scores for the two instruments are the same is described by the equation

$$S_{PCL} = a_0 + rS_{PCL5}, \quad (1)$$

where r is the ratio of scale ranges $(85 - 17)/(80 - 0) = 0.85$, and $a_0 = 17$ is the intercept. This is the line of percentage equality across instruments. However, a_0 and r can differ from expectation in practice because inventories have unique characteristics and distinct items. To visualize the similarity of data sources, we performed a combinatoric analysis of the 17 datasets which generated 153 dataset pairs. Six of these pairs are shown in Figure 2 as ranked paired sum scores subsampled from each dataset. Figure 2a compares three dataset pairs that used the same instrument, whereas Figure 2b contrasts three dataset pairs that used different instruments.

The deviation from the line of percentage equality (gray lines) is a convoluted measure of all the underlying differences between the two datasets. If all data were on this line, the two datasets would be identical and $R^2 = 1$. Overall, simple percentage conversions (gray lines) did not accurately predict raw scores across datasets in most cases, and percentile models also showed errors. These findings suggest corrections should be performed before linking across instruments.

Harmonization

The ComBat-GAM algorithm was implemented to correct for site-specific variations such as differences in severity-based inclusion criteria. Figure 3 shows all 153 R^2 comparison values grouped into boxplots by dataset. Intrinsic differences across datasets confound simple linear conversion as hypothesized (Figure 3a). After

harmonization (Figure 3b) the average R^2 increased by 0.74, and the number of well-fit pairs ($R^2 > 0.9$) increased significantly from 29 to 37 (+28%, $p < .001$).

Covariate Adjustment

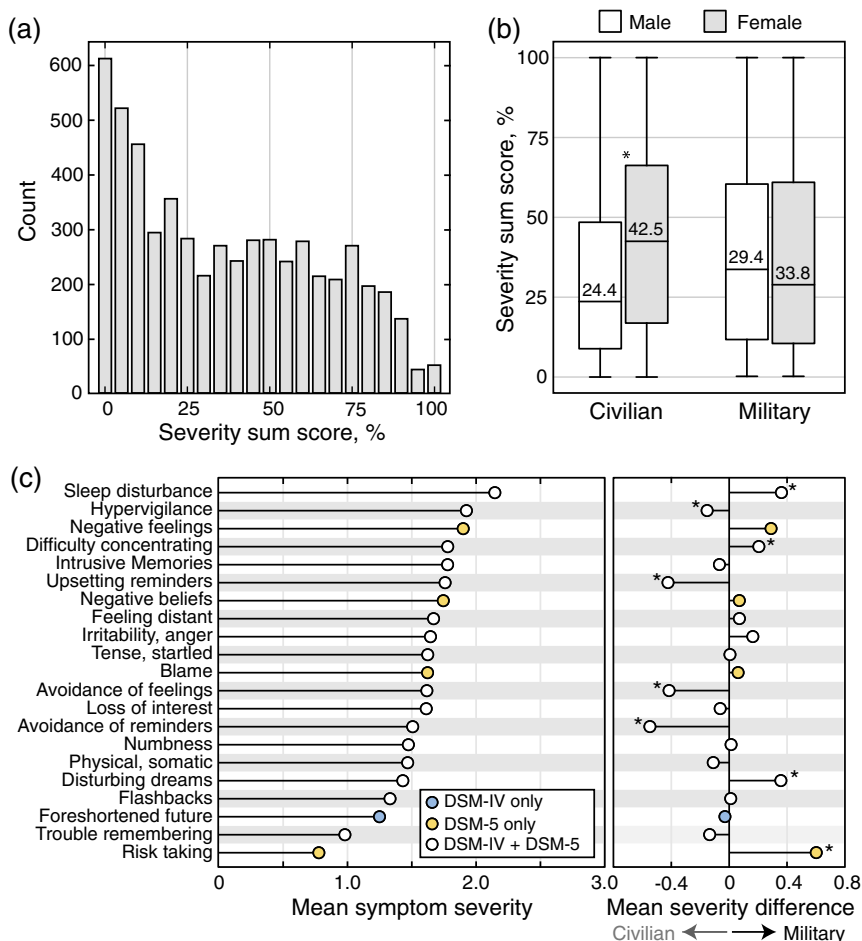
Table 2 shows the results for two blocked logistic regressions predicting symptom severity using (a) sociodemographic factors and (b) instrumentation. The models predicted postharmonized severity scores binarized to above/below the clinical cutoff of each inventory (e.g., $y = 1$ for scores $>32/80$ for PCL-5). The unharmonized model fits were poor, but improved after harmonization. After harmonization, military status showed much higher odds, $OR: 2.74$, 95% CI [2.38–3.16], of above-threshold severity, while age groups and sex showed broadly similar odds. Age, sex/gender, and population were adjusted out of the data using linear regression.

Converting Scores Across Instruments

After harmonization and covariate correction, we again consider the conversion of PCL-5 to PCL scores. To derive an empirical relationship between the instruments, we aggregated two datasets; one containing all postharmonized and covariate adjusted PCL-5 sum scores, and one containing all postharmonized and covariate adjusted PCL sum scores.

As described in the methods, we selected $N = \text{argmin}([n_a, n_b])/2 = 272$ samples from each group at random and sorted and paired the scores. A regression was performed to estimate r , the ratio of scale ranges for PCL-5 to PCL (nominally 0.85), and a_0 , the intercept (nominally 17). Over 10 random subsampling iterations, the regression found $a_0 = 19$, and $r_{alb} = 0.83$ was the best fit with a mean R^2 of 0.965. The similarity of the coefficients and the high R^2 of this

Figure 1
Summary of PTSD Symptom Reporting Across All Datasets



Note. (a) A histogram illustrates the distribution of all raw percentage severity sum scores. (b) Percentage severity sum scores are shown broken out as boxplots by military status and sex/gender. (c) A stem plot shows the mean severity reported for each item. (c, right) The average differences between military and civilian severity scores are shown for each item. PTSD = posttraumatic stress disorder; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders-5. See the online article for the color version of this figure. * indicates significance at $p < .05$ after q correction for multiple comparisons.

approach are encouraging, but nonlinearities could emerge for more distinct assessments. To account for this, a quadratic sum score term, S_A^2 , was added to the conversion models. Table 3 shows the estimated parameters for different instrument pairs using the following model:

$$S_B = a_0 + \beta_1 S_A + \beta_2 S_A^2, \quad (2)$$

where S_A and S_B are the sum scores of instruments A and B . This process was repeated for all “ $A \rightarrow B$ ” conversions.

Worked Example

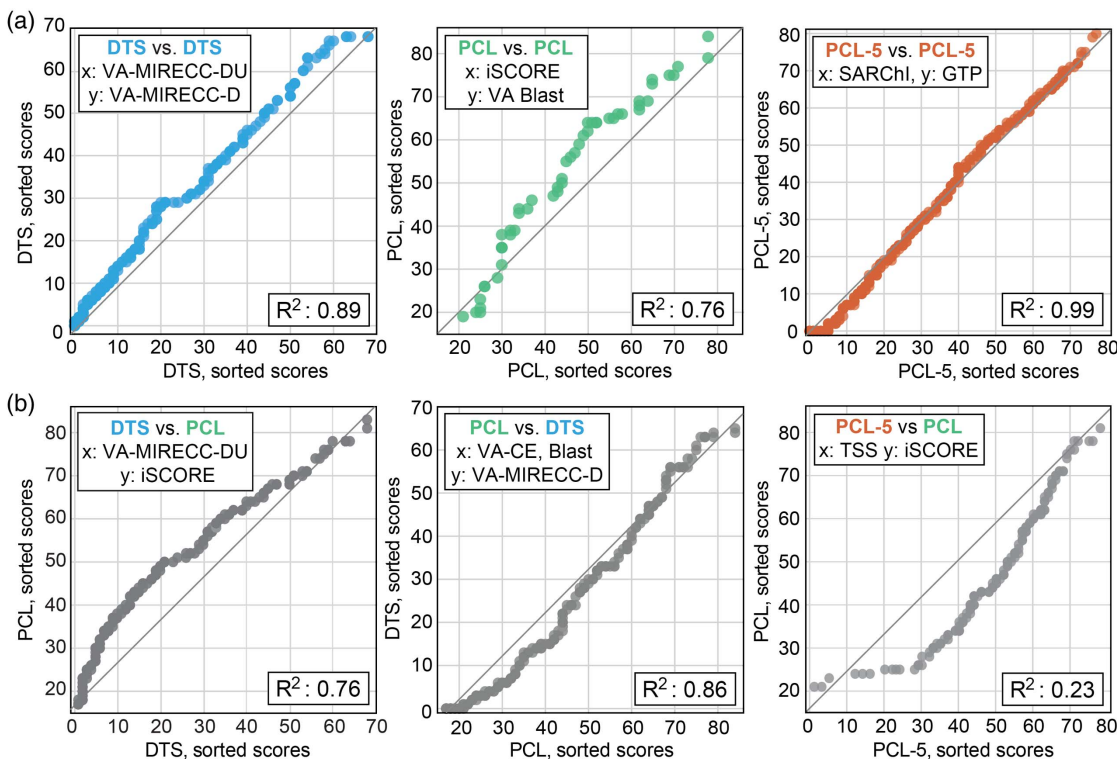
A worked conversion example is as follows: A clinician wishes to infer a patient’s PCL-5 score using the patient’s DTS score. Since

frequency scales were not considered in these models, the clinician calculates only the patient’s DTS severity score ($S_{DTS} = 45$ out of 68). The clinician consults Table 3 and finds the coefficients of $DTS \rightarrow PCL-5$ are $a_0 = -2.2$, $\beta_1 = 1.44$, and $\beta_2 = -0.0044$, with root mean squared error of $e = 1.5$. Using these parameters, S_{PCL5} is calculated as

$$S_{PCL5} = -2.2 + 1.44(45) - 0.0044(45^2) = 53.7 \pm 1.5. \quad (3)$$

To demonstrate conversion on real data, we elected to hold out an independent sample of $N = 1,212$ observations of the PCL-5 from LIMBIC-CENC (see Method section) not used in model training. The conversions from PCL-5 to other assessments are shown in Figure 4. Figure 4a plots the inferred severity scores as a function of

Figure 2
Visualizing Raw Severity Scores Across Datasets and Instruments



Note. (a) The sorted severity scores of datasets that used the same instrument are compared. The gray line indicates equality from the lowest to the highest possible score on each scale. The coefficients of determination (R^2 , inset) measure deviation from the line. (b) Like (a) but comparing datasets that used different PTSD assessment instruments. PTSD = posttraumatic stress disorder; DTS = Davidson Trauma Scale; iSCORE = imaging support for the study of cognitive rehabilitation; SARChi = Stellenbosch's South African research chairs initiative; GTP = Grady trauma project; PCL-5 = PTSD checklist for DSM-5; TSS = McLean trauma spectrum study; VA = Veterans Affairs. See the online article for the color version of this figure.

the input PCL-5 severity scores. Figure 4b shows a histogram of the severity scores after conversion to other severity scales.

Discussion

“Harmonization” is often used to refer to data aggregation, but true data harmonization aims to minimize source and measure variations in ways that preserve meaning. New PTSD treatments continue to be assessed (Rauch et al., 2019), but persistent gaps in clinical/research consistency, and shifting trial admission standards make it challenging to generalize findings. Harmonization presents a promising solution to address these concerns, and it is interesting to consider what new insights may be gleaned from data where source and acquisition effects are mitigated.

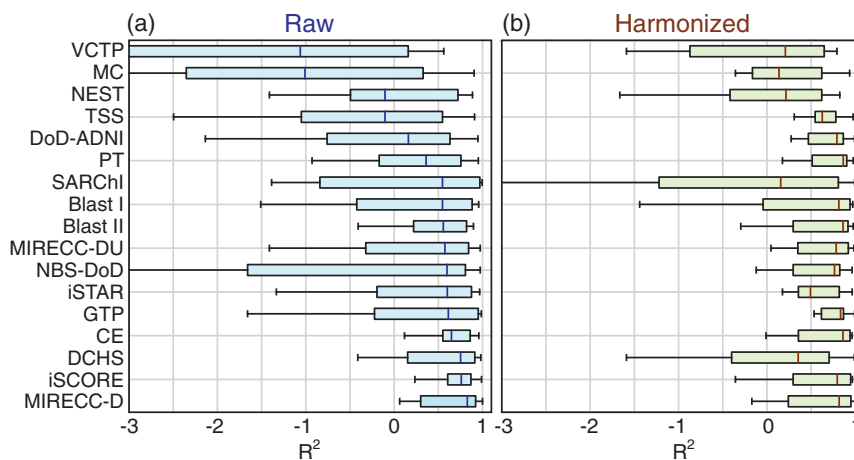
In the process of generating models for converting across PTSD severity scales, we found that empirical Bayes harmonization methods can isolate variations induced by different settings and procedures. We also confirmed our hypothesis that harmonization and covariate adjustments would significantly improve conversion model performance. There are many points of distinction that could explain why harmonization improves performance. For example, research studies and clinical facilities typically perform assessments in different ways. However, an exhaustive list of all the subtle ways

that sources may differ is not necessary in order for these effects to be empirically detected and removed in aggregate.

Similarly, we outlined a thorough description of all the ways that the instruments differed (see Method section), but ultimately, simple models effectively captured the instrumental variations without reference to their specific differences. Conceptually, these strategies draw from the observation that model explainability is distinct from predictive power (Kasirzadeh, 2021). Several facts suggest the instrumental components were well isolated. For example, simple percentage equivalence models fit well after corrections, but not before (Figure 2).

This work also provides some general insights into PTSD symptomology. If site, biological, and instrumental variations can be separated from individual symptom reports, then perhaps new intrinsic truths can be unearthed from previously confounded data. There are even hints of this possibility in this work, and while the influence of age and sex/gender variables were relatively unchanged by harmonization, military status exhibited a large increase in odds after accounting for site effects (from *OR*: 0.89 to *OR*: 2.74). These population specific differences were only apparent after harmonization, and similar approaches applied to problems in the behavioral sciences could help to identify hidden population effects.

Figure 3
Comparison of Pre/Post Harmonization Fit Quality



Note. (a) The distribution of coefficients of determination are shown as boxplots broken out by data source. (b) Like (a) but after Bayesian correction of site effects; VCTP = neuroimaging meditation therapy in Veterans with comorbid mild traumatic brain injury (TBI) and PTSD; MC = military mission connect study; NEST = McLean neurocardiac effects of stress and trauma study; TSS = McLean trauma spectrum study; PT = personality traits and brain matter aberrations as potential markers of mTBI and PTSD study; SARChI = Stellenbosch’s South African research chairs initiative; NBS-DoD = neural and behavioral sequelae of blastrelated traumatic brain injury; iSTAR = imaging study of trauma and resilience; GTP = Grady trauma project; CE = longitudinal study of chronic effects of TBI in Veterans and service members; DCHS = Drakenstein child health study of prenatal mothers; iSCORE = imaging support for the study of cognitive rehabilitation; MIRECC-DU = mental illness research education clinical, centers of excellence, Duke University; MIRECC-D = mental illness research education clinical, centers of excellence, Durham Veterans Affairs. See the online article for the color version of this figure.

Table 2
Results of Blocked Logistic Regression to Predict Elevated PTSD Symptom Severity

OR (95% CI)	Positive PTSD screen, Sum score > cutoff	
	Sociodemographic characteristics	
	Raw	Harmonized
	$R^2 = 0.007$	$R^2 = 0.091$
Age group (Ref: 18–29)		
30–39	1.18 (1.03–1.35)*	1.16 (1.0–1.34)*
40–49	1.14 (0.98–1.31)	1.01 (0.87–1.18)
50 and older	0.72 (0.60–0.85)*	0.7 (0.59–0.83)*
Sex/gender (Ref: Male)		
Female	1.16 (1.02–1.33)	1.06 (0.93–1.22)
Population (Ref: Civ.)		
Military	0.89 (0.58–0.77)*	2.74 (2.38–3.16)*
	Inventory	
	Raw	Harmonized
	$R^2 = 0.014$	$R^2 = 0.058$
Inventory (Ref: PCL-5)		
DTS	0.72 (0.63–0.81)*	2.51 (2.2–2.86)*
PCL	0.58 (0.48–0.72)*	1.84 (1.5–2.25)*
mPSS	1.25 (0.98–1.6)	1.18 (0.93–1.5)

Note. OR = odds ratio; CI = confidence interval; DTS = Davidson Trauma Scale; mPSS = modified PTSD Symptom Scale; PTSD = posttraumatic stress disorder; PCL-5 = PTSD checklist for DSM-5; Raw scores (left) and postharmonized scores (right) are shown for comparison.

* indicates significance at $p < .05$ after correction for multiple comparisons.

Table 3
Model Parameters for Converting PTSD Severity Scales

Conversion A → B	a_0	β_1	β_2	R^2	e (RMSE)
PCL-5 → PCL	19.43	0.572	0.0037	0.982	1.54
PCL → PCL-5	-33.5	1.82	-0.0065	0.983	1.65
PCL-5 → MPSS	-0.35	0.729	-0.0016	0.987	1.2
PCL → DTS	-21.1	1.157	-0.0012	0.971	1.76
PCL-5 → DTS	2.128	0.63	0.00322	0.992	1.23
PCL → MPSS	-22	1.252	-0.0053	0.976	1.42
DTS → PCL-5	-2.2	1.44	-0.0044	0.989	1.5
DTS → PCL	19.07	0.906	0.0008	0.971	1.73
DTS → MPSS	-3.14	1.088	-0.0058	0.992	1.06
MPSS → PCL-5	3.081	1.216	0.0077	0.986	1.54
MPSS → PCL	21.15	0.65	0.0158	0.975	1.66
MPSS → DTS	3.84	0.775	0.0135	0.993	1.18

Note. DTS = Davidson Trauma Scale; PTSD = posttraumatic stress disorder; RMSE = root-mean-square error; PCL-5 = PTSD checklist for DSM-5; MPSS = Modified PTSD Symptom Scale. a_0 = offset, β_1 = slope, β_2 = quadratic coefficient, R^2 = coefficient of determination, and e = root mean squared error.

Conclusion

We leveraged a multisite mega analysis ($n = 17$ datasets) to derive quantitative recommendations for the conversion of common PTSD severity scales. The data ensemble and the impact of site and covariates on severity scores were described. After isolating the instrumental component of severity scores, we produced accurate (median $R^2 = 0.985$) models for converting PTSD symptom severity scores, which were validated on data not used in model construction. This analysis suggests PTSD instrumentation has objective effects that can be isolated and removed, and these methods offer new empirical perspectives on the role of harmonization in the behavioral sciences.

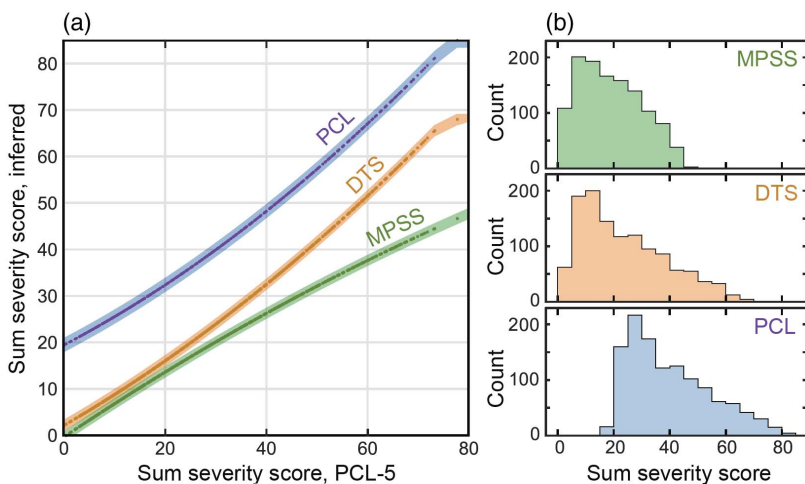
Limitations

We did not consider quantitation of structured interviews and future work could explore crosswalks between structured clinical interviews and brief inventories. PTSD is a highly comorbid condition, and the data collection across sources did not facilitate consideration of a wider range of conditions. We also did not consider varying trauma exposure or repeated measures, and future work would benefit from the analysis of multiple measures per person. This limitation is mitigated in part by repeated subsampling drawn from all sites in aggregate when deriving models. This work exclusively considered English language assessment and lacked data on race/ethnicity. Future analysis of assessment languages and race/ethnicity would be beneficial.

Individuals who reported the lowest possible sum score (e.g., a total severity score of zero on PCL-5) were excluded in order to remove predictive biases associated with differing inclusion criteria across studies. Some studies exclusively recruited individuals with moderate/severe PTSD severity, while others were convenience samples with a large fraction of participants with no history of PTSD exposure, who scored zero on PTSD severity assessment. Converted scores below zero do not provide additional value, and should be truncated to zero after conversion.

Our conversion models were constructed on data after site and covariate effects were removed. This means the models are only recommended for within-study inference unless additional harmonization procedures are enacted. In accordance with APA guidelines (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), we caution against broad generalizations of the methods presented here to new data and samples, or to cases where there is limited exact measure overlap. The extent to which harmonization and adjustment remove

Figure 4
Converting Severity Scores



Note. (a) Model-inferred PTSD severity scores on different instruments as a function of real PCL-5 severity scores. The shaded areas indicate ± 1 RMSE. (b) The equivalent PCL-5 severity score distributions after conversion. PTSD = posttraumatic stress disorder; RMSE = root mean squared error; PCL-5 = PTSD checklist for DSM-5; DTS = Davidson Trauma Scale. See the online article for the color version of this figure.

unwanted effects is empirical, and some residual effects may have persisted. We intentionally did not force the intercepts of the models, and at the extremes the parameters can return a value outside of the inferred inventory range. Any values out of range (e.g., negative numbers) should be truncated to the maximum/minimum possible value.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing: National council on measurement in education*. American Educational Research Association.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Boedhoe, P. S. W., Heymans, M. W., Schmaal, L., Abe, Y., Alonso, P., Ameis, S. H., Anticevic, A., Arnold, P. D., Batistuzzo, M. C., Benedetti, F., Beucke, J. C., Bollettini, I., Bose, A., Brem, S., Calvo, A., Calvo, R., Cheng, Y., Cho, K. I. K., Ciullo, V., ... the ENIGMA-OCD Working-Group. (2019). An empirical comparison of meta- and mega-analysis with data from the ENIGMA Obsessive-Compulsive Disorder Working Group. *Frontiers in Neuroinformatics*, 12, Article 102. <https://doi.org/10.3389/fninf.2018.00102>
- Brett, E. A., Spitzer, R. L., & Williams, J. B. (1988). DSM-III-R criteria for posttraumatic stress disorder. *The American Journal of Psychiatry*, 145(10), 1232–1236. <https://doi.org/10.1176/ajp.145.10.1232>
- Choi, S. W., Schalet, B., Cook, K. F., & Cella, D. (2014). Establishing a common metric for depressive symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychological Assessment*, 26(2), 513–527. <https://doi.org/10.1037/a0035768>
- Cifu, D. X., & Dixon, K. J. (2016). Chronic effects of neurotrauma consortium. *Brain Injury*, 30(12), 1397–1398. <https://doi.org/10.1080/02699052.2016.1219065>
- Davidson, J. R. T., Book, S. W., Colket, J. T., Tupler, L. A., Roth, S., David, D., Hertzberg, M., Mellman, T., Beckham, J. C., Smith, R. D., Davison, R. M., Katz, R., & Feldman, M. E. (1997). Assessment of a new self-rating scale for post-traumatic stress disorder. *Psychological Medicine*, 27(1), 153–160. <https://doi.org/10.1017/S0033291796004229>
- Donald, K. A., Hoogenhout, M., du Plooy, C. P., Wedderburn, C. J., Nhapi, R. T., Barnett, W., Hoffman, N., Malcolm-Smith, S., Zar, H. J., & Stein, D. J. (2018). Drakenstein Child Health Study (DCHS): Investigating determinants of early child development and cognition. *BMJ Paediatrics Open*, 2(1), Article e000282. <https://doi.org/10.1136/bmjpo-2018-000282>
- Falsetti, S. A., Resnick, H. S., Resick, P. A., & Kilpatrick, D. (1993). The Modified PTSD Symptom Scale: A brief self-report measure of posttraumatic stress disorder. *Behavior Therapist*, 16, 161–162.
- Gillespie, C. F., Bradley, B., Mercer, K., Smith, A. K., Conneely, K., Gapen, M., Weiss, T., Schwartz, A. C., Cubells, J. F., & Ressler, K. J. (2009). Trauma exposure and stress-related disorders in inner city primary care patients. *General Hospital Psychiatry*, 31(6), 505–514. <https://doi.org/10.1016/j.genhosppsych.2009.05.003>
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70–85. <https://doi.org/10.1177/2515245917751886>
- Kasirzadeh, A. (2021). *Reasons, values, stakeholders: A philosophical framework for explainable artificial intelligence*. arXiv:2103.00752
- Kennedy, E., Panahi, S., Stewart, I. J., Tate, D. F., Wilde, E. A., Kenney, K., Werner, J. K., Gill, J., Diaz-Arrastia, R., Amuan, M., Van Cott, A. C., & Pugh, M. J. (2022, February 5). Traumatic brain injury and early onset dementia in post 9-11 veterans. *Brain Injury*, 5, 1–8. <https://doi.org/10.1080/02699052.2022.2033846>
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602. <https://doi.org/10.1001/archpsyc.62.6.593>
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. Springer. <https://doi.org/10.1007/978-1-4757-4310-4>
- Lebois, L. A. M., Li, M., Baker, J. T., Wolff, J. D., Wang, D., Lambros, A. M., Grinspoon, E., Winternitz, S., Ren, J., Gönenç, A., Gruber, S. A., Ressler, K. J., Liu, H., & Kaufman, M. L. (2021). Large-scale functional brain network architecture changes associated with trauma-related dissociation. *The American Journal of Psychiatry*, 178(2), 165–173. <https://doi.org/10.1176/appi.ajp.2020.19060647>
- Logue, M. W., van Rooij, S. J. H., Dennis, E. L., Davis, S. L., Hayes, J. P., Stevens, J. S., Densmore, M., Haswell, C. C., Ipser, J., Koch, S. B. J., Korgaonkar, M., Lebois, L. A. M., Peverill, M., Baker, J. T., Boedhoe, P. S. W., Frijling, J. L., Gruber, S. A., Harpaz-Rotem, I., Jahanshad, N., ... Morey, R. A. (2018). Smaller hippocampal volume in posttraumatic stress disorder: A multisite ENIGMA-PGC study: Subcortical volumetry results from posttraumatic stress disorder consortia. *Biological Psychiatry*, 83(3), 244–253. <https://doi.org/10.1016/j.biopsych.2017.09.006>
- Norris, F. H., & Hamblen, J. L. (2004). Standardized self-report measures of civilian trauma and PTSD. In J. P. Wilson, T. M. Keane, & T. Martin (Eds.), *Assessing psychological trauma and PTSD* (pp. 63–102). Guilford Press.
- Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I. M., Satterthwaite, T. D., Fan, Y., Launer, L. J., Masters, C. L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S. C., Fripp, J., Koutsouleris, N., Wolf, D. H., ... Davatzikos, C. (2020). Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, 208, Article 116450. <https://doi.org/10.1016/j.neuroimage.2019.116450>
- Pugh, M. J., Kennedy, E., Prager, E. M., Humpherys, J., Dams-O'Connor, K., Hack, D., McCafferty, M. K., Wolfe, J., Yaffe, K., McCrea, M., Ferguson, A. R., Lancashire, L., Ghajar, J., & Lumba-Brown, A. (2021). Phenotyping the spectrum of traumatic brain injury: A review and pathway to standardization. *Journal of Neurotrauma*, 38, 3222–3234. <https://doi.org/10.1089/neu.2021.0059>
- Radua, J., Vieta, E., Shinohara, R., Kochunov, P., Quidé, Y., Green, M. J., Weickert, C. S., Weickert, T., Bruggemann, J., Kircher, T., Nenadić, I., Cairns, M. J., Seal, M., Schall, U., Henskens, F., Fullerton, J. M., Mowry, B., Pantelis, C., Lenroot, R., ... the ENIGMA Consortium collaborators. (2020). Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. *NeuroImage*, 218, Article 116956. <https://doi.org/10.1016/j.neuroimage.2020.116956>
- Rauch, S. A. M., Kim, H. M., Powell, C., Tuerk, P. W., Simon, N. M., Acierno, R., Allard, C. B., Norman, S. B., Venners, M. R., Rothbaum, B. O., Stein, M. B., Porter, K., Martis, B., King, A. P., Liberzon, I., Phan, K. L., & Hoge, C. W. (2019). Efficacy of prolonged exposure therapy, sertraline hydrochloride, and their combination among combat veterans with posttraumatic stress disorder: A randomized clinical trial. *JAMA Psychiatry*, 76(2), 117–126. <https://doi.org/10.1001/jamapsychiatry.2018.3412>
- Shelby, R. A., Golden-Kreutz, D. M., & Andersen, B. L. (2005). Mismatch of posttraumatic stress disorder (PTSD) symptoms and DSM-IV symptom clusters in a cancer sample: Exploratory factor analysis of the PTSD Checklist-Civilian Version. *Journal of Traumatic Stress*, 18(4), 347–357. <https://doi.org/10.1002/jts.20033>
- Suliman, S., Stein, D. J., & Seedat, S. (2014). Clinical and neuropsychological predictors of posttraumatic stress disorder. *Medicine*, 93(22), Article e113. <https://doi.org/10.1097/MD.0000000000000113>

- Sun, D., Gold, A. L., Swanson, C. A., Haswell, C. C., Brown, V. M., Stjepanovic, D., LaBar, K. S., Morey, R. A., & the VA Mid-Atlantic MIRECC Workgroup. (2020). Threat-induced anxiety during goal pursuit disrupts amygdala-prefrontal cortex connectivity in posttraumatic stress disorder. *Translational Psychiatry*, *10*(1), Article 61. <https://doi.org/10.1038/s41398-020-0739-4>
- Tate, D. F., Wade, B. S. C., Velez, C. S., Drennon, A. M., Bolzenius, J. D., Cooper, D. B., Kennedy, J. E., Reid, M. W., Bowles, A. O., Thompson, P. M., Gutman, B. A., Lewis, J. D., Ritter, J. L., York, G. E., & Bigler, E. D. (2019). Subcortical shape and neuropsychological function among U.S. service members with mild traumatic brain injury. *Brain Imaging and Behavior*, *13*(2), 377–388. <https://doi.org/10.1007/s11682-018-9854-8>
- Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolin, S., Bralten, J., Bruin, W. B., Bülow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C. G. F., Donohoe, G., Eyler, L. T., Faraone, S. V., Favre, P., . . . the ENIGMA Consortium. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Translational Psychiatry*, *10*, Article 100. <https://doi.org/10.1038/s41398-020-0705-1>
- Towse, J. N., Ellis, D. A., & Towse, A. S. (2021). Opening Pandora's box: Peeking inside Psychology's data sharing practices, and seven recommendations for change. *Behavioral Research*, *53*, 1455–1468. <https://doi.org/10.3758/s13428-020-01486-1>
- Weathers, F., Litz, B., Herman, D., Huska, J., & Keane, T. (1993). *The PTSD checklist: reliability, validity, and diagnostic utility* [Paper presentation]. In Annual Meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.
- Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., Keane, T. M., & Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, *30*, 383–395. <https://doi.org/10.1037/pas0000486>
- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD checklist for DSM-5 (PCL-5)*. National Center for PTSD. www.ptsd.va.gov
- Weiner, M. W., Harvey, D., Hayes, J., Landau, S. M., Aisen, P. S., Petersen, R. C., Tosun, D., Veitch, D. P., Jack, C. R., Jr., Decarli, C., Saykin, A. J., Grafman, J., Neylan, T. C., & the Department of Defense Alzheimer's Disease Neuroimaging Initiative. (2017). Effects of traumatic brain injury and posttraumatic stress disorder on development of Alzheimer's disease in Vietnam Veterans using the Alzheimer's Disease Neuroimaging Initiative: Preliminary Report. *Alzheimer's & Dementia*, *3*(2), 177–188. <https://doi.org/10.1016/j.trci.2017.02.005>
- Weis, C. N., Webb, E. K., Damiano, S., Larson, C. L., & deRoon-Cassini, T. A. (2022). Scoring the life events checklist: Comparison of three scoring methods. *Psychological Trauma: Theory, Research, Practice, and Policy*, *14*(4), 714–720. <https://doi.org/10.1037/tra0001049>
- Wilde, E. A., Dennis, E. L., & Tate, D. F. (2021). The ENIGMA Brain Injury working group: Approach, challenges, and potential benefits. *Brain Imaging and Behavior*, *15*(2), 465–474. <https://doi.org/10.1007/s11682-021-00450-7>
- Wilkins, K. C., Lang, A. J., & Norman, S. B. (2011). Synthesis of the psychometric properties of the PTSD checklist (PCL) military, civilian, and specific versions. *Depression and Anxiety*, *28*(7), 596–606. <https://doi.org/10.1002/da.20837>
- Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondanville, K. A., Hall-Clark, B., Foa, E. B., Young-McCaughan, S., Yarvis, J. S., Hembree, E. A., Mintz, J., Peterson, A. L., & Litz, B. T. (2016). Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment*, *28*(11), 1392–1403. <https://doi.org/10.1037/pas0000260>

Received October 28, 2021

Revision received March 31, 2022

Accepted April 3, 2022 ■